Titre: Title:	Dynamic pace control of e-lessons through time-scale modification
Auteur: Author:	Charles-Antoine Julien
Date:	2004
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Julien, C.-A. (2004). Dynamic pace control of e-lessons through time-scale modification [Master's thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7489/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/7489/
Directeurs de recherche: Advisors:	
Programme: Program:	Unspecified

UNIVERSITÉ DE MONTRÉAL

DYNAMIC PACE CONTROL OF E-LESSONS THROUGH TIME-SCALE MODIFICATION

CHARLES-ANTOINE JULIEN
DÉPARTEMENT DE MATHÉMATIQUES ET
DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTREAL

MÉMOIRE PRÉSENTÉ EN VUE DE
L'OBTENTION DU DIPLÔME DE MAÎTRISE
ÈS SCIENCE APPLIQUÉES
(GÉNIE INDUSTRIEL)
JUIN 2004

Your file Votre référence
ISBN: 0-612-97956-3
Our file Notre référence
ISBN: 0-612-97956-3

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

DYNAMIC PACE CONTROL OF E-LESSONS
THROUGH TIME-SCALE MODIFICATION

présenté par : JULIEN Charles-Antoine

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. BAÏLON Jean-Paul, D. Sc., président

M. ROBERT Jean-Marc, doctorat, membre et directeur de recherche

M. CYBIS Walter, doctorat, membre

ACKNOWLEDGEMENTS

I wish to thank Réal Héppelle and Pierre DeSchryver for their technical and hardware support during the production of the modified tutorial chapter. Their zeal and availability were greatly appreciated at a time when I felt rather lonely pushing a project which may never end. Although I know they probably do not see it that way I believe I owe them much.

Mention must be made of Jean-Paul Baïlon who allowed me to perform the empirical study with his course and tutorial. His approval given, I hounded the summer class professors Richard Lacroix and Abdel Kahalek Sami to get the study under way and collect the ensuing surveys. Their patience and support have proven critical to the completion of the empirical study.

I discussed my project with André Cincou between two laps in the pool and he proved to be of great value to get the empirical study underway. He made himself available to show me his online tutorial project and directed me towards other teachers who had educational technology interests. His interest boosted my morale when it wasn't clear how this project would end up.

Finally, I wish to thank Jean-Marc Robert for supervising this research and quickly correcting the final version of this memoir. I also appreciated his trust on the two occasions where he allowed me to perform some paid duties under his tutelage.

ABSTRACT

The benefits of digital multimedia interactive learning were heralded by the industry and accordingly, resources were generously allocated. Once the hype had abated, quantifiable and reproducible effects on learning efficiency were hard to come by. Debate over the role of technology in education is rare since educators are not trained in technology and technologists have little credibility with educators. Today, the vast majority of instruction is still performed in the same classroom context much as it has for the past centuries. What can the meta-medium of computing offer to the individual learner?

First, e-learning should make use of multi-modality, specifically audio narration augmented by static or dynamic visual material. Cognitive load theory supports the use of multiple information input channels as an effective way to reduce the risks overload. Secondly, as the text reader dynamically controls reading rate, the multimedia online student should be able to receive self-paced learning using time-scale modification (TSM). Using a slider control widget, the e-learner can speed up or slow down an audio/visual lesson without creating a "chipmunk" or "Mickey Mouse" effect. This research verified TSM benefits for online tutorials containing material judged technical.

An empirical study was performed using the first year engineered materials undergraduate class at l'École Polytechnique of Montréal. Live observations were conducted with five (5) voluntary students and 30 survey responses were collected. Results show TSM use creates an efficiency gain and is generally perceived as being useful. The empirical data suggests TSM generated a 30% time saving but, the most promising benefit is the sense of control given to the user. Indeed, 70% of survey respondents report dynamically modifying the pace of instruction according to content difficulty. Though this research concerns only novice use of TSM, based on the strong case built through a lengthy literature review and these promising empirical results, we strongly recommend real-time dynamic pace control of audio/visual lessons through TSM be given to the individual e-learner. This will be accomplished by supporting TSM integration with existing online content development software.

CONDENSÉ

Il y a eu une période d'engouement pour les technologies éducatives qui a accompagné l'augmentation rapide de la capacité interactive et multimodale des outils informatiques. Malgré les investissements significatifs, des résultats concrets et reproductibles concernant l'effet sur l'efficacité et l'efficience d'apprentissage sont difficiles à trouver. En effet, certains sont même prêts à affirmer, comme Richard E. Clark, que la technologie n'a pas plus d'influence sur l'apprentissage que le camion de livraison d'épicerie sur notre alimentation.

La littérature disponible dans le domaine de l'utilisation du multimédia en éducation est principalement constituée d'anecdotes et d'expériences ancrées dans un contexte spécifique. On se retrouve donc avec une grande quantité d'études de cas mais très peu de lignes directrices permettant de guider le design et ainsi, assurer une certaine qualité.

Comme la télévision durant les années '60, le méta-médium informatique fut publicisé comme outil de démocratisation de l'éducation et n'a pu répondre à ces attentes. En effet, les nouvelles technologies ont généralement besoin d'une période de maturation afin d'exploiter leurs qualités intrinsèques par rapport aux médias existants. Par exemple, le cinéma du début du siècle présentait simplement sur grand écran du théâtre filmé. Réciproquement, l'informatique éducative actuelle nous propose généralement ce que le papier et la télévision nous offrent depuis longtemps.

Il s'avère que quelques auteurs se sont intéressés spécifiquement à l'utilisation du multimédia en éducation. Principalement, Richard E. Mayer a développé un modèle d'apprentissage multimédia basé sur la théorie de la charge cognitive et il l'a testé dans de multiples expériences depuis plus de 15 ans. Entre autres choses, ses résultats et les connaissances cognitives acquises démontrent qu'un contenu de formation en ligne devrait, en général, être présenté à l'aide d'une narration audio au lieu d'une explication textuelle.

En effet, un texte écrit, que ce soit sur papier ou à l'écran, est d'abord un stimulus visuel et ensuite, une narration audio interne au lecteur. Dans le cas d'une formation en ligne, l'utilisation de représentations symboliques visuelles statiques (photos, illustrations, tableaux, etc.) ou dynamiques (animations, vidéo, etc.) accompagnées d'un texte explicatif tend à surcharger le canal visuel. Cet état résulte inévitablement en un filtrage cognitif et par le fait même, en une baisse de l'efficience d'apprentissage.

Des didacticiels audiovisuels sont disponibles sur le marché. Par contre ils permettent difficilement à l'apprenant d'effectuer des comportements d'acquisition d'information connus, comme par exemple le survol, les retours en arrière et les changements de rythme de lecture. Ces comportements cognitifs donnent au lecteur la capacité d'adapter le rythme de livraison du contenu en fonction de sa charge cognitive instantanée.

Comme la lecture, une formation audiovisuelle doit permettre à l'apprenant de modifier de façon dynamique le rythme de livraison de la matière. En effet, les travaux de David H. Reilly démontrent bien l'importance qu'a le rythme de livraison sur l'efficacité de l'apprentissage. À l'aide d'un modèle mathématique non linéaire, il a établi que l'optimisation de l'apprentissage demande un rythme de livraison proportionnel à la capacité cognitive de l'élève, et ce à tout instant. Ce rythme optimal doit stimuler suffisamment pour encourager une concentration maximale et ainsi réduire les distractions possibles. Par contre, il ne doit pas être trop élevé pour ainsi perdre l'étudiant en induisant une surcharge cognitive.

Ce rythme de livraison adéquat est nécessairement variable dans le temps pour chaque apprenant. La fatigue ou le stress auront de l'influence sur la capacité cognitive. Le niveau de connaissances antérieures dans le domaine et la familiarité avec la voix du narrateur auront également un impact sur le rythme de livraison optimale d'apprentissage.

Une leçon typique inclut habituellement des parties générales et descriptives qui peuvent être généralement accélérées. Dans certains domaines on retrouve aussi du contenu plus technique comme, par exemple, une démonstration mathématique d'un phénomène. Dans ce cas, on pourrait même s'attendre à une baisse du rythme de livraison par rapport à la normale.

L'utilité de la variation du rythme de présentation est connue par l'enseignant qui doit continuellement adapter son débit au groupe d'étudiants. Il espère que la plupart seront capables de suivre ou ne seront pas trop ennuyés. Cependant, il est inévitable qu'une partie sera perdue à cause d'un rythme trop rapide et qu'à l'inverse, une autre partie perdra un peu de son temps parce que le rythme sera trop lent.

On peut se demander pourquoi les formations en ligne avec narration audio n'offrent qu'un seul rythme de présentation. Certains diront que l'utilisateur de ces outils informatiques doit habituellement écouter un segment audiovisuel de quelques minutes et ensuite passer au prochain segment quand bon lui semble. Ce type d'interactivité de navigation n'offre pas un contrôle du rythme de livraison à l'apprenant. Il se compare plutôt à un livre où le lecteur peut à tout moment tourner les pages mais chacune doit systématiquement être lue au même rythme imposé.

De plus, la voix humaine est généralement émise à un rythme de 150 mots par minute (MPM) et les professionnels en communication peuvent se faire comprendre jusqu'à environ 175 MPM. En comparaison, un lecteur efficace peut lire jusqu'à 300 MPM et la grande majorité d'entre nous lisent entre 225 et 275 MPM. En effet, la langue parlée est parsemée de pauses, ralentissements et redondances. Ces chiffres nous forcent à nous demander pourquoi devrait-on perdre son temps à écouter une narration audio quand on peut être plus efficace à l'aide de la lecture. En acceptant qu'une formation audiovisuelle soit cognitivement plus efficace que l'équivalent en texte/visuel, que faire pour en augmenter l'efficience en MPM?

Il est possible de modifier dynamiquement la vitesse d'élocution d'une narration enregistrée sans en affecter la tonalité. La modification de l'échelle de temps (MET) se fait par un algorithme qui réduit la longueur d'un signal audio en y retranchant systématiquement des sections de 30 à 100 millisecondes. Les mots de la narration audio ainsi traitée sont perçus proportionnellement plus courts. Le signal audio peut alors être diffusé en moins de temps tout en restant compréhensible pour l'auditeur.

Il existe plusieurs algorithmes de MET mais les plus populaires et jugés plus pratiques sont ceux basés sur la méthode « overlap add » ou OLA. En effet, bien qu'en principe l'extraction de minces tranches de signal soit fort simple à effectuer, il en est autrement pour le raccord entre les tranches restantes. Un signal où les sections sont simplement raboutées une à l'autre comprend des distorsions et cliquetis désagréables. Afin d'améliorer la qualité des liens nouvellement créés, les algorithmes de type OLA tentent de joindre intelligemment les interfaces entres les tranches. On calcule l'endroit de ressemblance maximale entre les deux bouts de signal afin d'assurer, par exemple qu'un vallon ne soit directement juxtaposé à un sommet créant ainsi un signal nul.

La MET est connue depuis environ 40 ans et de nombreux auteurs de littérature scientifique en clament l'utilité pour une consommation efficace d'information audio. Les premiers appareils électroniques étaient fort dispendieux et par la suite les algorithmes demandaient trop de traitements informatiques. Depuis la fin des années '90, la puissance des processeurs disponibles sur tous les ordinateurs personnels a abattu les obstacles techniques à l'utilisation de la MET.

De nombreuses expériences scientifiques ont démontré que cet outil technologique peut procurer un gain d'efficience en éducation et ce, même pour un contenu jugé technique. On parle d'une accélération de 30 à 60% se traduisant, par exemple, en une économie de temps de 18 à 36 minutes dans le cas d'une leçon de 60 minutes. Après une courte période d'adaptation la plupart des utilisateurs choisissent un rythme de croisière entre 195 et 240 MPM et ne veulent habituellement plus retourner au rythme normal de 150 MPM.

De plus, deux études empiriques démontrent que la MET est grandement appréciée par les utilisateurs. Des étudiants universitaires et des employées de la compagnie Microsoft ont visualisé des conférences préenregistrées de 1.5 à 2 fois plus rapidement que normal. Le contenu était cependant de type introduction générale sans être technique. Outre ce gain d'efficience, cette fonction permettait aux étudiants de passer d'un mode de visionnement passif à un mode interactif où la vitesse de livraison pouvait être changée en quelques secondes.

Est-ce que ce gain d'efficience peut être vérifié dans un vrai contexte d'apprentissage et ce avec un contenu technique comprenant des démonstrations mathématiques, graphiques et modèles 3D? De plus, si la leçon n'est pas une conférence ou un cours magistral préenregistré mais bien un didacticiel spécialement conçu, y a-t-il autant de gains possibles?

Afin de répondre à ces questions, nous avons effectué une étude empirique auprès des étudiants du cours de sciences des matériaux de l'École Polytechnique de Montréal. Ce cours offre, en plus de séances magistrales et du manuel recommandé, un didacticiel d'apprentissage apprécié historiquement par 94% des utilisateurs. Les étudiants de la session d'été 2004 ont eu accès à une version du premier chapitre d'étude qui permettait le changement dynamique de l'échelle de temps à l'aide d'un contrôle produit par Enounce Inc. La durée normale d'écoute de ce chapitre est d'environs 35 minutes.

Après avoir utilisé ce chapitre test, trente (30) étudiants bénévoles ont rempli un questionnaire concernant leur utilisation de la fonction. L'analyse statistique démontre une réduction de 30% du temps de présentation moyen (min de 0,9 et max de 1,7 fois normale) et que 90% des sujets ont évalué le degré d'utilité de la MET à 4 et plus sur une échelle de 1 à 7. Les commentaires inscrits par les sujets nous indiquent qu'une minorité (5 sur 30) a fait une expérience d'apprentissage passive où la vitesse n'a jamais été modifiée en cours de leçon tandis que la majorité (21 sur 30) a interagit avec l'interface en fonction de la difficulté de contenu.

En effet, 70% (21 sur 30) des sujets affirment avoir dynamiquement modifié la vitesse selon la difficulté de la matière et seulement 17% (5 sur 30) indiquent qu'ils ont gardé un rythme constant du début à la fin. Ces comportements nous amènent à exprimer un bémol au sujet du gain de 30% d'efficience puisque la moyenne des vitesses les plus utilisées selon les sondages (1.28 fois normale) n'est probablement pas représentative de la vitesse moyenne effective. De plus, 33% (10 sur 30) des sujets ont réduit la vitesse sous la normale de 150 mots par minutes durant au moins un passage jugé plus compliqué ce qui devrait augmenter le temps d'instruction malgré le fait que les autres ont gardé un vitesse plus élevé que la normale.

Nous croyons que la simple utilisation de la fonction par une majorité des sujets est un endossement plus fiable pour la MET que le gain d'efficience de 30% rapporté. Puisque la densité et la complexité de l'information sont variables au cours de la leçon on ne peut affirmer que le rythme de présentation normalement imposé est optimal du début à la fin. En effet, certains sujets ont précisé qu'ils ont choisi une vitesse plus élevée durant les passages où seul un vidéo est présenté et plus près de la normale durant les passages contenant une démonstration mathématique, un graphique et une illustration dynamique du phénomène. Dans le premier cas, l'étudiant doit se concentrer sur une seule source d'information visuelle par contre, dans le second cas, on doit regarder et intégrer trois sources différentes ce qui augmente la charge cognitive et impose une réduction du débit de la voix pour prévenir une surcharge.

Nous avons également observé cinq (5) étudiants utilisant la MET dans le cadre de leur étude de la matière du cours. Ils ont accepté d'être observés pendant le visionnement du premier chapitre du didacticiel et de se faire interrompre à quelques reprises pour permettre à l'observateur de comprendre leurs actions. De ce groupe, quatre (4) sujets étaient en première ou deuxième année du baccalauréat en sciences appliquées et l'autre était au niveau du doctorat en sciences informatiques. Deux des sujets en étaient à leur deuxième tentative de réussite pour ce cours.

Les comportements observés étaient généralement très différents. Certains modifiaient souvent la vitesse selon la densité d'information visuelle de chaque segment, d'autres ont choisi une vitesse dès le début et n'en ont pas dérogé. Les vitesses utilisées ont varié de 0.8 à 1.6 fois la normale et tous le sujets observés ont affirmé que cette fonction était utile à l'apprentissage.

Il est intéressant de noter que 80% (4 sur 5) des sujets observés semblaient ralentir et accélérer le rythme de livraison de façon inversement proportionnelle à la densité d'information visuelle. Les vitesses rapides sont sélectionnées pour les segments contenant strictement un clip vidéo. Par contre, les segments montrant une démonstration mathématique ainsi qu'une animation ou un vidéo furent systématiquement ralentis puisqu'ils imposent une plus grande charge cognitive visuelle. Nous avons donc observé que plus la charge visuelle augmente plus le débit choisi de la voix est réduit.

En définitive, il nous semble que la MET est une technologie qui peut apporter des effets positifs à l'apprentissage en ligne même pour du contenu jugé technique. Cette conclusion n'est pas nouvelle et est supportée par une littérature scientifique considérable. La technologie est disponible et les bénéfices en éducation sont établis; alors qu'est-ce qui explique l'absence d'offre dans ce domaine?

L'étude des technologies éducatives est un domaine multidisciplinaire. Les éducateurs sont rarement des technologues et vice versa. Il y a donc peu de gens faisant la promotion scientifique de nouvelles possibilités technologiques en éducation. L'offre actuelle de contenu en formation en ligne reflète le constat que les éducateurs connaissent bien le design pédagogique basé sur une narration écrite. Les outils technologiques utiles à la consommation de narration audio sont peu connus dans une société à tradition écrite et seuls quelques auteurs scientifiques, Barry Arons de MIT par exemple, s'y sont intéressés.

Cet ouvrage aura donc servi, nous l'espérons, à promouvoir l'utilisation de la narration audio en formation en ligne à l'aide d'outils efficaces et adaptés à ce genre de livraison. Nous avons présenté un de ces outils, la MET, qui selon nous, doit être intégrée aux logiciels de conception de formation en ligne comme Macromedia Director par exemple. Les éditeurs de ce type de logiciel ne le feront que s'ils savent que leur clientèle le désire. L'étude empirique que nous avons produite créera peut-être une petite partie de cette demande.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LEXICON

CBT: Computer Based Training

CTML: Cognitive Theory of Multimedia Learning by R. E. Mayer.

TAS: Time-altered speech

TSM: Time-Scale Modification

VSP: Variable Speed Playback

TCS: Time-Compressed Speech

Auding: internal audio narration produced by silent reading

INTRODUCTION

This Masters memoir presents the results of over 12 months of the author's full time work. The initial motivation was to explore usage of digital multimedia in education and began with an in depth review of recent educational material design handbooks and manuals. We narrowed our research into the usage of audio in e-learning and related audio information consumption tools. It became clear that in a society based on the written word, tools allowing efficient acquisition of audio information were promoted by few scientific authors and little known to the general public.

In the majority of cases, online educational content draws solely upon the visual channel through written text, static visual symbols (photos, diagrams, graphs, etc.) and dynamic visual support (video, animation, etc.). Since most educated adults are proficient readers and audio production costs are much higher than written text alone this situation is not surprising. In this context, one could argue the digital medium has basically reproduced what the printing press has been offering since the 17^{th} century. Indeed, if we ask the student to read characters on screen, he or she will inevitably print the content and the computer is reduced to a simple information delivery channel like traditional mail.

If the digital medium is to play a role in education it must leverage its distinguishing characteristics in order to leverage a competitive edge over traditional educational media like print and television. We will argue this set of qualities includes multi-modality, interactivity and dynamic display. Using these three criteria, the digital medium can surpass the traditional educational media offer.

There is considerable scientific support for the cognitive efficiency provided by the usage of both the audio and visual channels during information transmittal. When audio narration is used instead of text along with supporting static and dynamic visual material we reduce the probability of visual channel overload. This entails one should generally design an online lesson using audio narration as the main content delivery method instead of the written word.

In addition to the cognitive aspects, audio narration provides a myriad of additional benefits heralded by the scientific community. For example, a human tough is added to the lesson, a clear agency is established as the voice becomes a uniquely identifiable audio persona and variation in speed and energy can accentuate certain passages. Some authors are willing to state audio narration is simply better and more natural than written text. Indeed, one could argue we all first learned to hear and listen before we became readers.

In conjunction, education psychology shows the pace of instruction is a critical factor in learning effectiveness. Matching the rate of content delivery with the individual learner's capacity to absorb information is necessarily beneficial to the learning process. In the typical classroom a teacher must adopt an acceptable rate of speech so most students can follow the narrative line. This necessarily entails some will find this "middle ground" rate too fast and become lost through cognitive overload while others may be under-stimulated and lose concentration.

The one-to-one relationship between the content and the online learner permits an individualized and dynamic rate of presentation as dictated by the user in real-time. This type of functionality should allow the lesson's audio narration and related visual support to be instantly sped up or slowed down through an interface widget without loss of signal quality.

To this affect, time-scale modification is a tried and tested signal processing algorithm which effectively shortens or lengthens a speech signal without modifying the tone of the voice. The result is an audio narration whose words are heard proportionally quicker or longer without a "Mickey Mouse" or "chipmunk" effect associated with increasing the playback speed of an audio recording.

Although there is much educational TSM experimental support and two recent empirical studies, we have not found empirical data demonstrating benefits with an online tutorial presenting content judged technical in nature. Additionally, if usage in online education is so promising, why is TSM not widely known and available? Potential answers to these questions are the main value proposition of this research.

CHAPTER 1: KNOWLEDGE ACQUISITION

The following chapter presents three learning and knowledge taxonomies in order to express the scope of the current research. A brief description of constructivism is presented and finally, we present a simplified generic model of learning referred to in later chapters.

1.1. Learning taxonomies and constructivism

There are several taxonomies describing the learning process. Bloom (1969), Palkiewic (1986) and Rummelhart (1988) respectively list 7 competences, 4 types of knowledge and 3 types of learning. This educative technology research is concerned with the subset of the learning process in which the learner is asked to initially acquire domain vocabulary and, subsequent review purposes. Table 1.1 presents this scope within the above knowledge acquisition taxonomies delimited by bold characters.

Table 1.1. Learning taxonomy scope of current research

Bloom (1969) (Competences)	**Palkiewic (1986)** (Knowledge Types)	**Rumelhart & Norman (1988)** (Types of learning)
Motor	Decisional	**Accretion**
Knowledge	Rational	Tuning
Comprehension	Intellectual	Restructuring
Application	**Empirical**	
Analysis		
Synthesis		
Evaluation		

In fact, little is understood concerning the cognitive process of learning and comprehension. During the '80s, constructivism replaced behaviourism as the accepted model of the human learning process. This model states that units or chunks of information are received by the student through the sensory channels (usually audio and visual). Learning is performed by storing these chunks in long term memory. This potentially creates a path between new knowledge and the person's established relational web of concepts, facts, procedure, etc. This simplistic description is obviously not complete but,

> "while there are many interpretations of constructivism, the description provided by Soloway et al. as learning and understanding being 'active, constructive, generative processes such as assimilation, augmentation, and self-reorganisation' (p.190) captures the essence of the constructivist perspective" (Squires and Preece 1999, p.468).

Notably, constructivism supported a return to "teaching by storytelling" transmitted to current generation students of education. For example, The "Power of Story" (Collins and Cooper 1997) is a mandatory reference at the McGill University Education Faculty in Montreal.

1.2. Simplified learning stages

Since "it is helpful to model learning as a progression through stages" (Mayes and Fowler 1999, p.487) we have created a summarized learning process resulting in three major stages. Figure 1.1 illustrates this process.

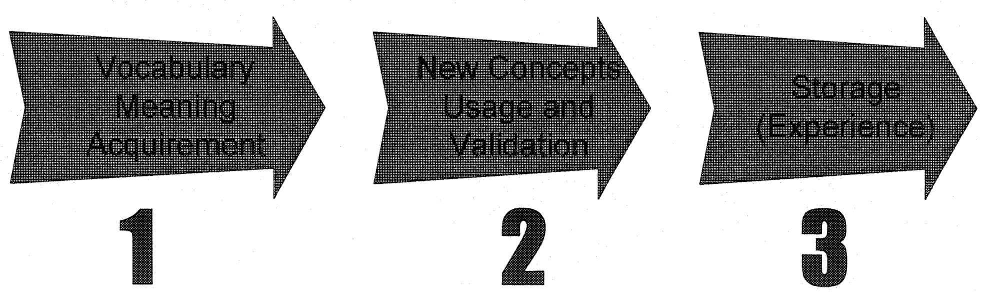

Figure 1.1. Simplified knowledge acquisition steps

Stage 1 shown in the above figure can be equated to the research scope defined in Table 1.1. Relative time spent by the individual learner within each of these stages can be debated since there is no predictive model of knowledge acquisition. One could get caught up for some time in any of these stages but eventually acquire the desired concepts or skills. Looping between two stages is highly probable.

Mayes et al. (1999) refer to this first stage as the *primary exposition* and state "the most important single function of the primary exposition is to orient the learner towards the subject matter" (p.493). This initially provides the learner with a map of the knowledge to be acquired and the related set of definitions.

The digital medium should be able to provide added efficiency tools for tasks within one or more stages. Specifically, we believe the first stage, initial vocabulary acquirement, is especially promising for an audio/visual medium which could offer a memorable overview of the domain. Concept manipulation and experimentation, we believe, requires more support in the form of what is traditionally called "coaching". Although Lucy Suchman believed this to be the "promise of interactive computer systems" (Suchman 1987, p.19) research into intelligent tutoring systems (Cerri, Gouardères and Paraguaçu 2002 for example) shows the difficulties associated with this avenue.

In order to discover the current state of affairs concerning digital tools in education the following chapter presents a review of recent e-learning literature. This will lead us to the scientific rationale for using multi-modality to reduce cognitive load. Finally, supported by scientific literature, it will be argued that as reading allows learner pace control over content delivery so should computer based tutorials.

CHAPTER 2: MULTIMEDIA IN EDUCATION

This section presents a picture of the state of multimedia use in e-learning design. It seems that, like other media innovations of the past, digital training material has reproduced what existing educational support have long been offering. As well, we now have empirical data describing e-learners as a special breed of students with their particular behaviours. Using the computer and internet as a means to learn collaboratively has also generated a massive amount of literature. Finally, we attempt to establish a basis on which digital tools can complement and ultimately compete with current educational media such as paper and television.

2.1. State of affairs

A review of literature concerning digital multi-media usage in education reveals the role of digital technology is being questioned. One of the most staggering blows to the e-learning panacea was given in 1983 by Richard E. Clark when he wrote:

> "The best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievements any more that the truck that delivers our groceries causes changes in our nutrition" (Clark 2001b, p.2).

This catchy phrase threw a wrench in the multi-media learning hype of the mid-eighties. Clark's assertion is that learning has to do with pedagogical design rather than means of delivery.

Substantiating the lack of multi-media learning support, in a well documented educational media volume, "The Knowledge Medium", Gary A. Berg observes "the use of media does not yet have strong theoretical and scholarly support" (Berg 2003, p.110). He acknowledges scientific authors like Clark and observes that "often in educational technology literature, the transmission of knowledge through media is seen as neutral— (in his opinion) this is far from the case" (Berg 2003, p.112).

2.2. Repurposing older mediums

Exploring current e-learning initiatives, one quickly observes that text is the most widely used and recognized educational medium. Indeed, "the classic information bank in the classroom is, of course, the text (as the) source of explicit information" (Perkins 1992a, p.46).

Most computer based content relies on text and visual support (static or dynamic). Although audio narrated digital content is becoming more widespread (Dreamweaver training from Macromedia for example), "data reveal(s) that those responsible for developing distance learning courses are still primarily focused on text-based approaches" (Berg 2003, p.106).

A text based e-lesson does not generally offer a significant advantage over paper medium except for efficient document distribution over the Internet. This is not a new phenomenon in communication media. Indeed,

> "Marshall McLuhan (1964) pointed out that new media are dependent on old media until the unique features of the new media are appreciated and developed." (Berg 2003, p.114) and,

> "Repurposing goes hand in hand with the birth of any new medium. Film reused play…" (Negroponte 1995, p.63).

Generally, computer based educational tools have reproduced what paper or television have long been offering. Indeed

> "Skeptics might point out that the Web initially provided only a fairly static information distribution channel and served simply to off-load the task of photocopying from the institution to the individual" (da Graça Pimentel, Ishiguro, Kerimbaev, Abowd and Guzdial 2001, p.354).

One does see innovative uses of interactive 3D models and practical exercises. These utilize the digital medium's unique interactive and visual qualities but at some point, meaning must be associated with manipulation. This sense making narrative is generally transmitted by on-screen text. If one is going to read textual characters we believe paper serves as better support medium.

2.3. Multimedia instruction

Multimedia instruction is not a digital creation.

> "The classroom itself can be seen as a rich multimedia environment where dense audiovisual information is combined with miscellaneous annotating activities that support the teaching and learning experience" (da Graça Pimentel, Ishiguro, Kerimbaev, Abowd and Guzdial 2001, p.357).

What is new and potentially beneficial to learning are the qualities the digital medium provides to the modern educational system. Since computers and their software are tools that allow one to make tools, it can be difficult to precisely define what constitutes their role in education.

In an effort to define CBT content, the following lists the characteristics of a "Multimedia instructional systems" (Clark 2001c, p.226):

a) Information rich displays where almost all is visual...
b) The potential for high levels of real time interactivity...
c) Maximum learner control of instructional access, pacing, scheduling, feedback, structure...

Clark referred to current educational multi-media as mostly visual since on screen text is the most popular narrative artefact. Although these characteristics can foster much discussion when considering learning impact, we note maximum learner control of pacing is mentioned.

Learner control is made possible by the one-to-one relationship between the learner and the computer. This agency should allow the student to acquire information when, where and at his or her own pace. Indeed, "empowering learning environments, then, are those in which options are rich, and dependent upon the response of learners" (Malone and Lepper 1987, p.238). This differs from the conventional classroom setting where the teacher dictates the pace of instruction for all.

Although personal agency influences learner control of instruction, it should be used with care. The learner cannot be allowed to blindly surf through an e-lesson. The correct sequence of concepts and relations should be maintained in accordance with the lesson objectives. One "must take care to maintain a narrative, in collaboration with the user" (Laurillard, Stratfold, Luckin, Plowman and Taylor 2000, p.17).

Narration is the art of telling a story. The voice one hears when reading these words. A well designed narrative allows the receptor to gradually construct a mental representation of the content. Since the "predominant opinion (…) is that vision dominates audio", the narrative is often transmitted textually. This satisfies the reader looking for entertainment but for the student "adding pictures to auditory or printed information enhances learning" (Jasper 1991, p.167).

Designers of any product are asked to consider the behavioural trends of the target users. This should also be the case for the development of computer-supported learning environments. The following section presents recently reported e-learner behaviour found in the scientific literature.

2.4. E-learner behaviour

Following an empirical study involving 9000 users sponsored by Microsoft Research, the ensuing scientific publication described online training sessions where "57% are shorter than 5 minutes, and only 17% are longer that half an hour" (He, Grudin and Gupta 2000, p.130). The authors conclude from these numbers that online "viewers are restless" (p.131). This is not surprising since there is virtually no social enforcement for training attendance and assiduity included within the one on one online learning context. They conclude e-learners are generally "browsing and skimming more than viewing (or reading)" (p.132).

We believe it would be considered unwise to ignore this known behaviour. Tools should be provided to quickly browse, scan and navigate through online content. This is not a revolutionary concept since printed text inherently offers these types of capabilities and we believe audio-visual digital content should as well.

2.5. Collaborative Learning

There is extensive empirical data on collaborative learning. The British Open University for example, has collected and published much practical experience. Gilly Salmon (Salmon 2002) shows high numbers of students and numerous repetitive design iterations to support a method of online collaborative learning using newsgroups, e-mail and chat support. In this case, collaboration is used primarily for synchronous or asynchronous communication between the learner and the class (newsgroups), the learner and a topic discussion group (live chat with 5 students and a tutor) and e-mail to class or team members.

We are less inclined to delve into this subject since one cannot reconstruct a course from the mountains of data contained within all e-mail, newsgroups and chat logs. Conversation, live or not, is inherently unstructured and difficult to interpret. Before any discussion takes place, some accepted base of domain vocabulary must exist between the participants. This dictionary of terms and associations must be learned as is, and verified at some later date through practical exercises, discussions with peers and experience. With the right tools, we believe this *spoon feeding* of initial domain vocabulary could efficiently be performed via digital medium.

2.6. E-learning's Edge

What is the role of computing technology in education? Is it another attempt to mass distribute knowledge as television in the 60's promised and failed to do or, does the meta-medium of computing give some advantage to the learner? Can it be a useful tool to render learning more efficient as it has with other types of cognitive tasks like word processing for example?

Since most computer based learning is currently text based, we believe the learner inevitably prints the material. In this strictly visual learning environment, paper serves as a more portable, tactile and recognized medium. Purely visual CBT cannot compete with this cheaper and static communication support. What unique advantages can digital technology offer the learner?

The literature hints at various unique qualities of the meta-medium of computing which could tip the balance in its favour versus paper. For example, "Gibbons and Fairweather (1998) identify five attributes that make the computer unique as an instructional medium: a dynamic display, the ability to accept student input, speed, ability to select, and flawless memory" (Berg 2003, p.115). On this last point, Shneiderman (1999) states computing "gives the ability to have a perfect record of learning" cited in (Berg 2003, p.115).

To simplify we have grouped "ability to accept input" and "ability to select" under the concept of interactivity. Additionally, we believe multi-modality capabilities (principally audio/visual) are also an important digital characteristic.

Comparing with other media, paper offers strictly visual stimuli and television offers no interactivity. Table 2.1 presents an evaluation of each medium according to the criteria of interactivity, multi-modality and dynamic display.

Table 2.1. Medium comparison under computing unique set of characteristics

Criteria	Education material medium			
	Computing	**Paper**	**Audio**	**Audio/Video**
Interactivity	High	Medium	Low	Low
Multi-modality	Medium	Low	Null	Medium
Dynamic display	High	Null	Null	High

Using this strategically chosen set of criteria, computing is the clear leader with audio/video a close second. On the other hand, paper would score highest if considering portability and laptops would provide a medium score to computing.

Concerning multi-modality, only audio and visual outputs are readily available on current computers. This explains why computing and audio/video are both judged as providing medium multi-modality capabilities. Uses of tactile feedback could increase this assessment in favour of the digital medium. For example, when planning a road construction project, force feedback through a mouse or joystick while the user hovers over a population density map can allow one to feel the "path of least resistance".

We have established that if computing is to have a competitive offer in education it should be through the use of some of its unique set of characteristics such as interactivity and multimodality. Learning is inherently a cognitive process and digital tools can ultimately only offer the promise of increased efficiency. The following chapter presents the cognitive rationale for the use of multimodality in educational media.

CHAPTER 3: THE CASE FOR MULTIMODALITY

This chapter is based upon the works of Richard E. Mayer who has been experimentally researching cognitive impacts of multimedia in knowledge acquisition. We then explain the relation between reading and audio narration and finally, an inventory control analogy is used to enhance the reader's understanding of the function of working memory.

3.1. Cognitive theory of multimedia learning

In order to evaluate multi-media's effect on learning one requires a model representing the cognitive process involved. Stage 1 of the learning process illustrated in Figure 1.1 is classically performed by reading and listening to lectures. In recent years, video and computer based lessons have been produced in order to support this task. As we have seen in previous sections, CBT is still predominantly a text based narration and static or dynamic visual cues.

Cognitive load theory provides a dual channel model based on Pavio's (1986) dual coding theory and Baddeley's (1998) theory of working memory. Richard E. Mayer presents the cognitive theory of multimedia learning (CTML) illustrated in the simplified version shown in Figure 3.1.

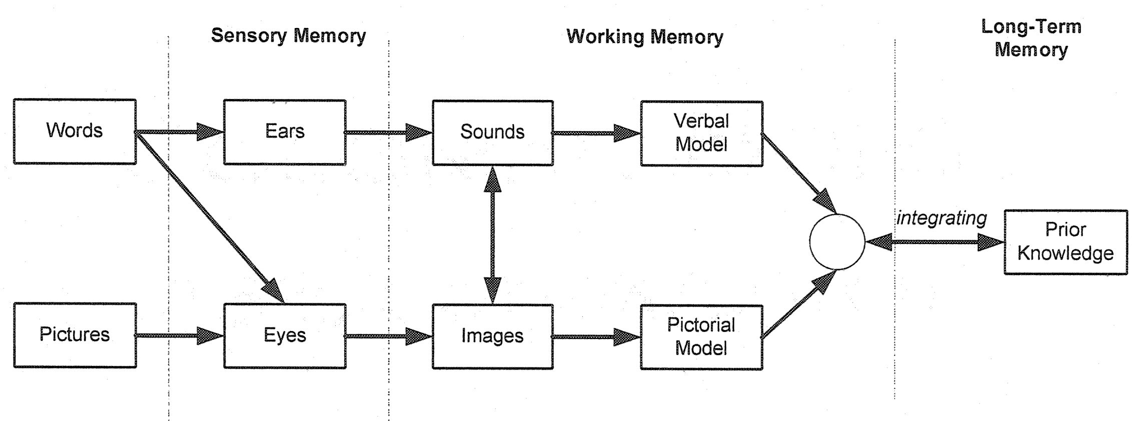

Figure 3.1. Cognitive theory of multimedia learning (Mayer and Moreno 2003)

We have modified the aero linking "Prior Knowledge" and the preceding integration point to reflect flow in both directions assuming prior knowledge is eventually modified by working memory content. The original model showed a one-way flow from long term memory to the integration point.

For the past 10 years, based on this model, Mayer and his colleagues have conducted over 100 experimental tests leading to 10 research-based principles for how to design on-line learning environments (Mayer 2003a). The model has held its grounds during experiments concerning text placement and content redundancy for example. In addition to North-American acknowledgements, international support is given to Mayer's work by Dubois et al. (2000) from the Université Pierre Mendès in Grenoble, France.

He suggests channel overloading as the first type of learning induced cognitive overload. When text and visual cues of any kind (photos, graphs, animation, etc.) are shown together the result is a *split-attention effect* ((Sweller 1999) cited in (Mayer and Moreno 2003)) where the visual working memory becomes saturated by characters and pictures. The ensuing bottleneck engenders a filtering of subsequent information. Indeed, "the eyes receive a lot of concurrent information but only some of that information can be selected for further processing in visual working memory" (Mayer and Moreno 2003, p.32).

This phenomenon was also observed during experiments performed at New South Wales University. They conclude that "inclusion of written material, because it involved visual presentation, apparently interfered with the visualization needed to recall the message" (Mousavi, Low and Sweller 1995, p.321).

3.2. Reading as listening to one's self

Jones (1986) exposes the notion of reading as being visual, then auditory. This cognitive process is accomplished first through the visual sensory channel where word symbols are temporarily stored and interpreted, then through the auditory channel where they are internally heard as an audio narrative from which the user attempts to make sense in light of current knowledge. Jones refers to this internal narration as "auding". Figure 3.2 illustrates this process.

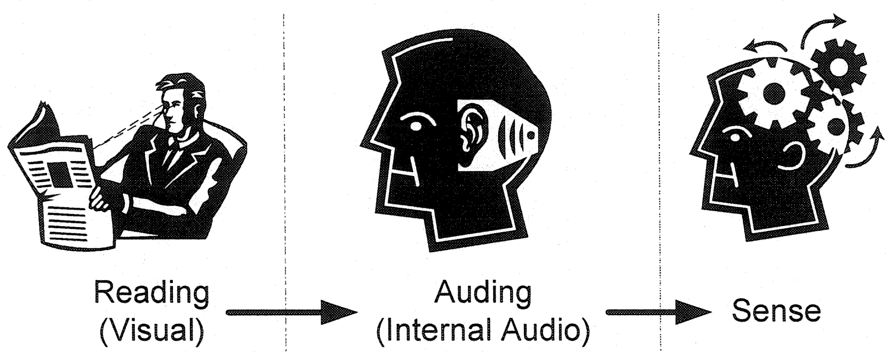

Figure 3.2. Reading as a visual and audio process

Alternatively, when one listens to audio speech the visual channel is left free to perform other tasks since only the audio channel is used to transmit the narrative line as illustrated in Figure 3.3.

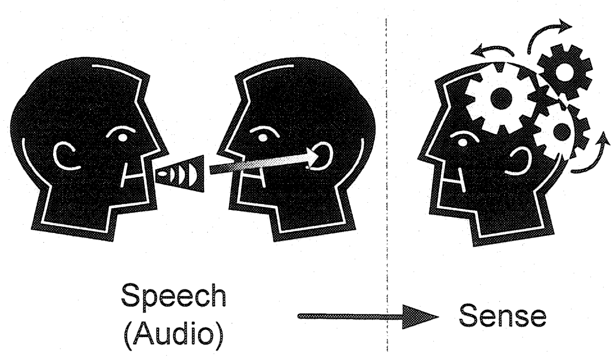

Figure 3.3. Process of listening to audio speech

In the case of conversation, the eyes continuously interpret the talkers' body language and lip movements for example. The previous two figures suggest listening should be cognitively less demanding than reading since no visual work is required to maintain the meaning narrative.

Reading as a visual then an audio cognitive process explains why "training in listening skill transfers to reading skill" (Jones 1986, p.53). A better listener is also a better "auder".

Reading as a visual then an audio cognitive process explains why "training in listening skill transfers to reading skill" (Jones 1986, p.53). A better listener is also a better "auder". Substantiating this notion, experiments have shown that "children who read well also listen well" (Hildyard and Olsen 1982, p.30).

Additionally, "the image is often an abstraction" (Jasper 1991, p.167) where a diagram for example, has only an indirect relation with its real-world counterpart. In these cases as Jasper states, narration is not the main point but supports understanding of the symbolic visual cue. The visual and associated narrative (text or speech) information do not provide redundancy and must be integrated to extricate meaning. When narration is text based, the learner's visual channel is forced to interpret the textual characters and temporarily store the resulting narration while attempting to integrate the auded result with the image.

The following section will attempt to illustrate what is understood to be the practical cognitive effect when both text and pictures are presented simultaneously by using an inventory management analogy known as a storage buffer.

3.3. Working memory buffer

As we have seen, each channel (audio and visual) possesses a distinct working memory (WM) which, upon reaching its capacity must filter additional input, creating a buffer effect. During integration with prior knowledge, WM space is made available for new input.

Stuart K. Card provides the following WM definition:

> "Working memory functions as (1) a place to hold operands, things to be operated on by the operations of cognitive processing; (2) a cache to hold in a rapidly accessible state recently input or used information; and (3) a buffer between processes that happen at incommensurate rates." (Card 1989, p.203).

He summarizes the WM definition by stating that "working memory is simply the set of all cognitive units in declarative memory activated at some point in time" (p.209).

When reading, the eye is forced to fill the visual buffer with characters and extricate meaning. The packets are held in visual WM and share the space with other related visual information (graphs, pictures, etc.). By definition a buffer has a fixed capacity and can only accept new input by liberating space. Consequently, "it is warranted to assume that channel capacity of learners is an important variable in the learning process (Britton & Tesser, 1982; Embretson, 1995; Huber, 1993)" cited in (Reilly 1998, p. 155).

As a result, it is believed "learners who experience less cognitive load are better able to mentally organize the presented material into a cause-and-effect chain and to mentally relate the material with relevant prior knowledge." (Mayer and Chandler 2001, p.395). Less should not be interpreted as null. There is a level of cognitive demand where the student is challenged by the material but possesses cognitive capacity to analyse and integrate the new concepts into long-term memory.

Jones (1986) refers to this point as the *comprehension asymptote*. This represents the level of WM demand where the learner has reached his or her individual level of acceptable ambiguity. The resulting pace of delivery varies for each individual, in time and depending on prior domain knowledge. Additionally, the cognitive load induced by a specific lesson is also expected to vary. Thus, one could say the information input capacity relating to a student's comprehension asymptote is equivalent to the WM processing cycle time.

3.3.1. WM capacity optimization through pace control

Cognitive load theory leads us to believe WM management is critical to the process of learning. Too much demand and the learner is lost and disoriented in a sea of meaningless words. Adversely, low demand from one specific source frees the WM for environmental elements not relevant to the task and concentration staggers, the student becomes distracted. "The benefits of using technology in the classroom may be lost if the amount of information presented is either too little or too much, or too fast or too slow for the learner to process effectively" (Reilly 1998, p.163).

The importance of the pace of instruction matching the learner's capacity to process the information is supported by constructivist literature. "It is the job of the constructivist teacher (or interactive technology) to help learners in their 'zone of proximal development'" (Perkins 1992b, p.163). This *zone of proximal development* is described as the point where the learner is provided with just enough information to keep the lesson challenging but not overwhelming. This concept is also used by Vigotsky (1978) and is analogous to the comprehension asymptote mentioned by Jones (1986).

The student, as the principal actor in the learning process, is considered the first and most accurate evaluator of personal cognitive load. A learner with the functionality to dynamically vary the rate of delivery should be able to better optimize his or her WM buffer utilization.

Illustrating this concept, variable rate of content delivery is performed intuitively by attentive teachers in a classroom setting. Information is presented in accordance to the perceived average group information absorption capacity. The resulting speed of delivery is inevitably slow for some, fast for others, but acceptable for most students. The lecturer evaluates capacity by asking questions, observing body language, listening and his/her interpretation skills.

3.3.2. Increasing WM capacity through multi-modality

Cognitive science literature states one may increase the comprehension threshold through multi-modality. Indeed,

> "if, as suggested by cognitive load theory, working memory limitations must be a primary consideration when instruction is designed, then we have a theoretical rationale for the use of dual-mode instructional techniques" (Mousavi, Low and Sweller 1995, p.321).

Based on their experimental results, they state that "a mixed auditory and visual mode of presenting information was more effective than a single (visual) mode" (p.323) and conclude "effective working memory capacity may be enlarged by using multiple channels" (p.319).

> (b) working memory capacity can be increased with a dual-presentation mode" (p.332).

Replacing text with an audio narration forces the learner to ignore irrelevant environmental audio stimuli. Indeed, attention is more concentrated when one is listening to an audio signal. There is a "greater dependence on short-term memory for listening, as compared to other language skills" (Parshall 1999, p. 7) because, for example "attribution of correct meaning often involves 'reading between the lines' during listening. Correct interpretation (...) requires careful attention to more than words" (King and Behnke 1989, p.431). When one is listening to audio speech, the mind is less free to wander as compared to reading the same narrative.

As shown previously in Figure 3.2 and Figure 3.3, using an audio narrative instead of text frees visual channel capacity. Additionally, integration of both channels may involve unique meta-cognitive processes where "in dual mode complementary processing takes place, the student using each mode to its advantage" (Jasper 1991, p. 166). Indeed, it

> "appears that when integration of simultaneously presented auditory and visual material is required, a broader, more specialized cognitive process of integration is utilized that is different from the process used in processing either auditory or visual information" (Fisher & Rubenstein, 1975) cited in (Reilly 1998, p. 162).

As well, people tend to have a stronger modality where they may favour either the visual or audio channel or, even have mixed modalities. Relative to age, Aarntzen (1993) observes the "proportion of persons with mixed modalities is larger among adults than children" (Aarntzen 1993, p.360). This would indicate usage of both audio and visual channels is especially fitting when the e-learners are expected to be of adult age as is often the case in distance education.

Finally, "the evidence of enhanced working memory capacity under dual-mode conditions is very strong. (...but) they have never been applied in a classroom context" (Mousavi, Low and Sweller 1995, p.321). We would argue a teacher continuously operates in audio/visual mode since students are usually expected to listen while watching. At a minimum, we would especially fitting when the e-learners are expected to be of adult age as is often the case in distance education.

Finally, "the evidence of enhanced working memory capacity under dual-mode conditions is very strong. (...but) they have never been applied in a classroom context" (Mousavi, Low and Sweller 1995, p.321). We would argue a teacher continuously operates in audio/visual mode since students are usually expected to listen while watching. At a minimum, we would expect digital learning environment to take advantage of this established and well practiced learner sensory behaviour. Providing only visual content in a digital learning environment is indeed a poor substitute for the multi-sensory experience of a live speaker.

This chapter has laid the foundations to support multimodality usage in e-learning content design. Practically, this translates into an increased importance for audio information delivery. In support, the following chapter presents scientific literature substantiates audio usage in computing and online training content.

CHAPTER 4: THE CASE FOR AUDIO

The following chapter first provides recommendations as to why and when one should use audio narration as opposed to text when developing e-content. A small sample of the extensive literary support for audio usage in computing and e-learning is then presented as well as audio lesson design guidelines. Finally, it is asserted that efficient audio information delivery requires specific tools not currently readily available.

4.1. Usage guidelines

As we have presented in section 3.3.2, multi-modality can be used as a way to increase the comprehension threshold by reducing visual channel demand. Mayer and Moreno (2003) suggest "the use of narrated animation represents a method for offloading (or reassigning) some of the processing demands from the visual to the verbal (audio) channel" (Mayer and Moreno 2003, p.32).

Specifying the context in which audio narration should be encouraged, they "recommend that you put words in spoken form rather than printed form whenever the graphics or animation is the focus of the words and both are presented simultaneously" (Clark and Mayer 2003, p.86).

A certain amount of text, a few key words, can be tolerated when they are integrated with the visual content (Mayer and Moreno 2003, p.37). When the focus is expected to be on the graphics and the narrative is continuous, one should generally be spoken to.

This seems reasonable since humans constantly process data from multiple senses and perform very well at it. We are adept at interpreting concurrent multi-channel input. We wonder if there is any reason not to offer multi-modality in digital learning environments other than absolute cost considerations.

Accordingly, we are interested in the capacities of the auditory channel and its impact on computer supported learning. The following section builds a case for the use of audio in computing and online learning.

4.2. Support for audio in computing

Sound carries unique qualities providing additional richness to the learning experience. Some have touted this untapped wealth. For example, "embedded in sound (human voice, audio narration) is information about feelings" (Negroponte 1995, p.148) and, there are "nuances carried by the audio signal but not in the words themselves" (Hindus, Schmandt and Horner 1993, p.378).

We intuitively assess relative importance of different parts of an audio narration based on the voice inflections and energy variations. Indeed "spoken language has a number of cues just as accentuation and pace, that pictures (or text) don't have" (Jasper 1991, p.171). The learner listening to audio narrated speech can "hear the speaker's emphasis and inflections upon what was important" (Schwab and DeGroot 1993, p.182) thus providing an implicit summary of the content.

The work of Barry Arons concerning speech skimming and audio navigation is, among others, based on the assertion that

> "speech is a very rich communications medium, layers of meaning can be embedded in intonation that cannot be adequately captured by text alone. Listening to speech is 'easier' than reading text – it takes less effort to listen to lecture than to read a paper on the same subject" (Arons 1991, p.143).

The relative cognitive effort required to listen as opposed to reading is not quantified. Generally, we intuitively agree with this statement based on the fact that humans naturally learn listening much earlier than reading and, as seen in section 3.2 (Reading as listening to one's self), reading demands an extraneous visual effort.

Additionally, the motivational aspects of the lesson delivery method are also influenced by different modalities. Fulford (1992) states that "unlike text, audio provides a human touch to instruction (...) the voice can motivate the learner through the enthusiasm expressed for the subject" (Fulford 1992, p.1).

Some take a strong stance by simply stating that "narration is better than writing" (Bruning, Horn and PytlikZillig 2003, p.12). We do not necessarily believe one delivery method is better than the other and this research attempts to establish potential benefits of the auditory channel and the settings where it may be favoured.

Concerning the contexts in which audio narration should be advantageous, after experimental results demonstrated that "the Sound group showed a greater percentage of correct information than statements from participants in the Text group" (Mann 1997, p.139), Mann (1997) states "in most settings, the preference for assessing modality effects in instruction should be verbal information" (p.134).

Reinforcing audio narrated speech value, "listening comprehension is, in general, superior to reading comprehension" (Jones 1986, p.56) and, he observed statistically that "reading comprehension is more variable than listening comprehension" (p.57).

Additionally, after experiments comparing reading versus listening in children, one finds "listeners may tend to recall more of the gist of the story and readers recall more of the surface verbatim features of the story" (Hildyard and Olsen 1982, p.20). This would indicate that when the pedagogical objective is to acquire a mental model of a particular concept, this should be performed through audio narration. On the other hand, learning a set of disconnected words would be more efficiently performed through written text. In support, Goldstein (1940) states that

> "listening comprehension (is) superior to reading comprehension" and, supporting Jones' previous assessment, "listening comprehension is less variable than reading comprehension" cited in (Olson and Berry 1982, p.2).

Finally, "the combination (of audio and visual) is a favourable one because pictures support general descriptive knowledge and audio text supports detailed specification" Nugent (1982) cited in (Jasper 1991, p.170).

Strictly Audio lessons can have value in certain contexts but, as we have seen in section 3.3.2, it is believed multi-modality benefits learning. Indeed, "the number of items people can remember is larger if they can simultaneously make use of several modalities (visual, motor, auditory)" (Baddeley and Hitch, 1974) cited in (Card 1989, p.207). Kozma (2001) adds that "while audio may be sufficient for those knowledgeable of a domain, visual symbol systems supply important situational information for those less knowledgeable" (Kozma 2001, p.154).

At this point, we believe we have built a strong case for working memory's influence on learning capacity and positive effects of multi-modality, more specifically audio and visual modes. The digital medium's capabilities in this area are well touted but, how does one go about designing, from the start, a lesson based on audio narrative instead of the printed word? The following section will attempt to give some practical answers to this question.

4.3. Building a lesson based on audio narration

Accepting the value of audio narration along with visual content, the lesson should be designed around a digital voice recorded audio script. There are professional speakers who lend their voices to audio-personas for commercial interactive phone systems for example. The persona's speech inflections, accents and energy variations are perceived and remembered by the listener. The persona becomes a unique auditory personality.

Concerning the characteristics of the voice, "speech should be exceptionally clear and preferably augmented with graphics" (Aarntzen 1993, p.359). On the other hand, for the visually impaired, "speech quality is less important than speed, which can and should be much faster than normal speech" (Anderson-Inman et al. 1990) cited in (Aarntzen 1993, p. 360). We believe the voice used should be one of a professional or semi-professional speaker able to control and fine tune his or her articulation, energy and tones.

The speaker should adopt an informal or conversational speech style since

> "research on discourse processing shows that people work harder to understand material when they feel they are in a conversation with a partner rather than simply

> receiving information (Beck, McKeown, Sandora, Kucan, and Worthy, 1996)" (Clark and Mayer 2003, p.136).

Accordingly, concerning the design of computer lessons, they recommend one "include some spoken or printed text that is conversational rather than formal (with clear agency)" (p.134). For example, the formal statement:

"The following section explains the use of audio narration in e-learning" can be informally reworded to:

"I will attempt to explain to you the use of audio narration in e-learning so you may easily understand."

There is an obvious difference between the formal speech styles heard during a lecture as opposed to being coached individually by a tutor. The one to one relationship between user and computer makes online learning closer to the latter.

This is in stark contrast to the formal and serious tone lecturers may adopt. Such style of speech is said to be aimed to no one in particular or to the group in general. A more conversational style of speaking reaches out to the individual listener and personally includes the learner in the lesson. As mentioned earlier, the learner attaches a personality to the recorded voice and agency is established. An endorsement of agency's value is also given by Brenda Laurel who believes computing interface designers should "make clear who is speaking…represent (the) source of agency" (Laurel 1993, p.142). We also believe online learning should be personally engaging, even for adults.

The possibilities of the human voice can also be specifically utilized. For example, "Reigeluth (1987) mentions raising the voice as one of the possibilities to gain attention" (Aarntzen 1993, p.361) and "the use of variety of tone, pitch, and volume in speech can be of help" (p.361).

The material designer can build an audio narrative supported by related visual cues. This multi-modal lesson demands listening concentration from the learner and allows visual rest when no visual stimulus is provided. Throughout the audio narration, a gentle tone can warn the user of visual content appearance and departure. The length of this tone can signal relative time of the visual input. We recommend a black screen when no visual support is provided.

Static and dynamic visual supports are to be inserted in association with the audio narration time-line. Figure 4.1 presents an example of the audio and visual time lines of a fictitious lesson.

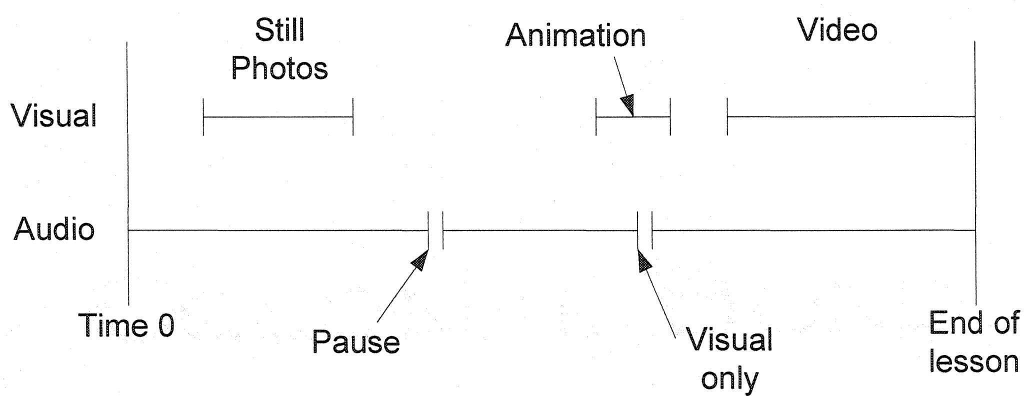

Figure 4.1. Synchronized individual channel time lines of an audio based lesson.

Computer-aided education should incorporate dynamic content when it is useful. Indeed, the use of animation can be, among others, to show transition between two states (Baeker and Small 1990) or simulation and related manipulation supports learner exploration and practice.

As well, "showing the total duration of a recording, along with a continually updated position indication during playback, is important for navigation and user comfort" (Myers 1985) cited in (Hindus, Schmandt and Horner 1993, p.388).

4.4. Audio skimming and navigation

Printed training material has the advantage of being widely recognized as a communication medium. The educational system makes us all practice reading from an early age. A paper document can be visually scanned for relevancy, relative size can be physically weighed, readers can regress, loop, visually spot known words to analyse if the contents are new or unknown, etc. For example,

> "readers with considerable domain knowledge would sometimes progress through the text at a rapid rate, using a single word to skip a vast amount of information (…) readers encountering difficulties (…) slow their rate and regress over passages" Bazerman (1985) cited (Kozma 2001, p.145).

In discussions with educated adults, we have found resistance to audio narration instead of the equivalent written material. They are usually proficient readers and possess cognitive abilities to quickly ingest vast quantities of text. We assume they believe reading is more efficient since "speech is naturally slow to listen to, and difficult to skim" (Arons 1997, p.26). Indeed, in his often cited document, Junor (1992) observed that the "most vexatious characteristics of tape-recorded information is the difficulty of locating a desired section rapidly and with ease" (Junor 1992, p.94).

With these audio speech drawbacks in mind, Barry Arons, initially at MIT, set out to design and collect software tools to provide audio skimming and navigation capabilities in order to make it " easier and more efficient to consume recorded speech (and) making one's ears as usable as one's eyes as a means for accessing stored information" (Arons 1997, p.26).

Of the many audio skimming and navigational tools he devised and assembled, the use of time-scale modification (TSM) was found to be a strong method to render audio more efficient. TSM is presented in detail in CHAPTER 6:.

At this point, we believe this research has established that if computer based training is to compete with other media, its content design should be based on an audio narrative instead of text and provide interactive audio skimming and navigational functionalities.

Notably, these should allow the learner to accelerate or decelerate the rate of audio narrative delivery much like you as the reader modify your reading rate depending on factors such as prior domain knowledge and concentration capabilities. The following section will establish the case for student controlled rate of content delivery in e-learning.

CHAPTER 5: THE CASE FOR PACE CONTROL

As we have seen, a multi-media digital lesson in which the main narrative is provided by audio narration instead of text is thought to be less cognitively demanding. On the other hand, since most adult learners have become very efficient readers, digital audio learning material competes poorly with the print medium.

This chapter describes established efficient reading behaviours and proposes that an audio narrated e-lesson should offer the same possibilities. Based on chaos theory, a model of learning is then presented demonstrating the importance of pace control over content delivery for efficient knowledge acquisition. We then refocus pace control's usefulness within the simplified learning process stages shown in CHAPTER 1: (KNOWLEDGE ACQUISITION) and finally, the measurable effects of pace control in e-learning are explained mathematically.

5.1. Pace control of content delivery

Print possesses a unique functionality no other medium currently offers: user controlled rate of content delivery. The reader can read at whatever rate he or she desires and indeed, influenced by cognitive load, readings rates often vary in time. Given the choice between an audio narrated, visually augmented lesson and the equivalent print material, one may still favour paper if only for its scanning capabilities.

As well, when one is reading a text the level of domain expertise will allow a different pace of content absorption. Indeed,

> "on the one hand, the novice will need a lot of help and therefore the learning material should be divided into small steps. On the other hand, the expert does not want to experience inadmissible waiting and therefore wants to skip parts of the waiting and therefore wants to skip parts of the lesson material and wants to go through the programme in much bigger steps than the inexperienced learner" (Aarntzen 1993, p.359).

Thus, if one is to produce an audio-visual lesson, it should incorporate tools to effectively scan and skip content like printed text inherently provides.

The pace of instruction is thought to be an important factor in learning effectiveness. Reilly (1998) states "there is a high degree of support in the literature for assuming that the rate of instruction is a critical factor in the instructional process" (Reilly 1998, p.155).

Paradoxically, originally published in 1991, Kozma (2001) states that "surprisingly little research has been done on the effect of pace on the process of learning with television and other transient media" (p.155). A live lecture for example, is a transient media since it does not allow the listener to dynamically backtrack and review emitted content. The only non-transient media currently available is text. It has its advantages but, current public education reliance on live lecturing would indicate transient media has strong support. An audio narration with or without visual content, is inherently transient in nature.

5.2. A non-linear model of learning

In order to evaluate the influence of different variables on learning effectiveness, David H. Reilly developed a model of learning based on non-linear system theory or theory of chaos. Table 5.1 presents this model[1].

Table 5.1. Learning Effectiveness Equation (Reilly 1998)

Variables		
LE	=	Learning Effectiveness (%)
R	=	Items per Inst. Period (#)
Te	=	Teacher Effectiveness (%)
Le	=	Learner Effectiveness (%)
IK	=	Initial Knowledge Base (%)
Reilly equation		
$LE = R * Te * Le * IK * (1 - IK)$		

[1] The variable names have been modified to facilitate understanding. The model is otherwise identical to the original.

IK was set at 0,001 to represent no initial knowledge and, although it is agreed Te and Le can be highly variable, he chose to set these at constant values since "it is the rate of instructional delivery that can most be controlled" (Reilly 1998, p.160).

He went on to test the behaviour of the model through a large number of teaching sessions or iterations and concluded that the rate of delivery "must be controlled to match the learner's ability to effectively process information" (p.160) and "the rate of presentation is a critical variable in determining the amount that will be learned" (p.160).

A graph was produced showing the impact of different R values on LE. He observes that an "r between 6.0 and 7.0 resulted in the highest, sustained Learning Level" (p.160). We could interpret this number as units, chunks, concepts, relations etc. per instructional period. This number coincides with accepted WM capacity since "Waldrop (1987) reports that the capacity of short term memory is no greater than about seven meaningful chucks of data" cited in (Reilly 1998, p.160).

Reilly finally suggests "obviously, matching the rate of presented information to learner's capabilities will assist them in optimizing the amount learned" (p.161) but, does not offer a way to quantify the added value of this "assistance" to the learner. This research attempts to quantify the usefulness and user appreciation of pace control in e-learning.

The motivation of the student is also affected by pace of content delivery. Indeed, research has shown "people prefer an optimal level of challenge" (Malone and Lepper 1987, p.231). Optimal in this case means not too easy and not too difficult. One must take for granted this level is highly variable and can only be known by the learner at any point in time.

Taking pace control as a critical learning variable, audio narrated lessons do not yet compete with paper. Indeed "print material provide greater pacing flexibility and may therefore offer a learning advantage over audio presentations" Reiser and Gagné (1983) cited in (Aarntzen 1993, p.358) and, print material's flexibility is also used to scan, backtrack, reread, etc.

Table 5.2. Comparison of learner pace control capabilities of different media

	Digital audio narration	Paper	Recorded Audio	Recorded Audio/Video
Pace control	Low	High	Low	Low

As mentioned earlier, it is a fact that readers instinctively vary their rate of content intake by varying the number of words read per minute. This is a natural and cognitively efficient way to manage informational input. Indeed,

> "it is (...) desirable to fill the cognitive capacity with the learning stimulus at an optimum rate so that distractive thoughts do not have time to enter the working memory" (Fulford 1992 p.2).

When the rate of delivery is too slow the "result is a listener who becomes bored or whose attention begins to wander" (Olson 1985, p.3).

Measuring the rate of delivery is generally performed by counting the number of words per minute the learner absorbs. This is simple enough to understand for an audio recording. In the case of a printed text, one can ask the student to read a complete passage and divide the total number of words by the number of minutes used to complete the text. The following equation presents this simple operation:

$$RateOfDelivery(WPM) = \frac{TotalWordsInText}{NumberOfMinutesToReadText}$$

In support, "Foulke and Sticht (1966) state that it is best to specify the wpm as this method is the most appropriate and is sufficient" (Harrigan 1996, p.9).

There is anecdotal support for user controlled pace of delivery. For example, during computer based audio instruction, "several of the students' written and verbal remarks (noted they were) 'waiting' for the audio to finish" (Barron and Kysilka 1993, p.286).

Finally, besides efficient WM management, pace control allows online students to save time by speeding up the audio narration and provides skimming functionality. Indeed, "the true (learning) bottleneck to address is the limited time people have" (He and Gupta 2001, p.382).

5.3. Range of promising return

One could argue other forms of assistance like direct teacher support for example, may be more effective to regulate the pace of delivery. We are not suggesting e-content pace control will replace the teacher. It is hoped however, that the teacher's role may be less one of a lecturer and more one of coach. Once the lesson narrative is audio-visually transmitted to the individual student at his or her own dynamically chosen pace, the teacher's role is to support exploration and untangle misunderstandings.

We believe e-lessons allowing dynamic pace control show greatest promise to significantly assist the learner during the first phase of the learning process. Subsequent uses would be for reference purposes. Useful knowledge will inevitably remain freshly available through repeated practical exercises. Figure 5.1 illustrates this range within the simplified learning process shown in CHAPTER 1:.

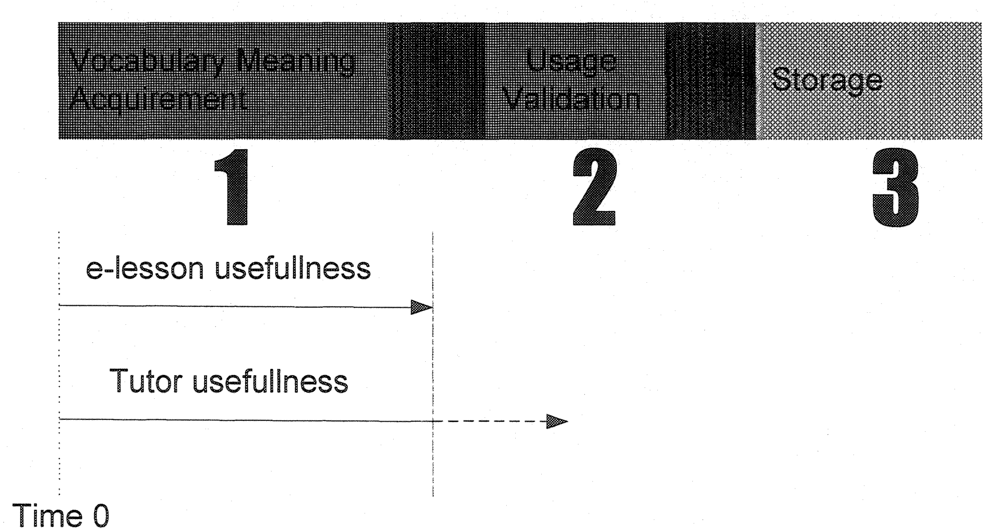

Figure 5.1. E-lesson vs. tutor usefulness in learning process.

As shown in the above figure, we do not profess to replace the human tutor but possibly reduce its need during the initial domain vocabulary acquirement stage. This may produce an effect where "the role of teacher will change to a manager and facilitator of learning" (Squires and Preece 1999, p.470).

When speech is highly accelerated, only previously known words are recognized by the listener. One can thus determine if the audio information is known or requires further in depth listening to be understood. This type of functionality supports the learning process since "adjustment of the level of description of learning material to match the moment-to-moment needs of the individual learner is one of the primary goals of tutoring" (Mayes and Fowler 1999, p.493).

5.4. The effect of pace control

As stated previously, we wish to quantify pace control's effect on learning efficiency. Efficiency is described as the number of knowledge units learned divided by instructional time. Fulford (1992) provides learning efficiency definitions shown in Table 5.3.

Table 5.3. Learning Efficiency Definition according to Fulford (1992)

Fulford Learning Efficiency Definition
Effectiveness: number of objectives mastered Efficiency: effectiveness divided by instructional time multiplied by 100.
Thus, $Efficiency = \frac{\#ObjectivesMastered}{InstructionalTime} \times 100$

Inspecting the learning efficiency equation it follows that a constant number of objectives mastered and a reduced instructional time increases efficiency. Practically speaking, if a given lesson is absorbed by a learner in less time, learning efficiency increases.

It may be that total instructional time will not decrease since a learner may slow down dense passages below normal and effectively cancel the time saved over sped up passages.

Although measurable learning efficiency would be null in this case, the fact that the student actually varied the pace of delivery would indicate pace control was considered useful learning.

Can efficiency gains be produced by dynamic pace control of CBT delivery? This type of interactivity with the interface would allow one to speed-up or slow down the audio narration measured in words per minutes (WPM). It goes without saying that related visual content must be synchronized with the accelerated or decelerated audio narration. How does one dynamically vary the WPM rate of an audio speech signal? The following chapter presents time-scale modification as a tool to provide such functionality.

CHAPTER 6: AUDIO NARRATION PACE CONTROL

This chapter proposes time-scale modification (TSM) as a viable technology to provide user pace control of audio information delivery. For comparison purposes, known reading and speech rates are presented and limiting factors to increased listening speed are described. There is a reported relation between upper reading and listening rates which is explored and the effects of TSM training are illustrated by relevant scientific literature. Section 6.6 then describes in layman's term the signal treatment preformed by TSM on a speech recording and since we are dealing with audio/visual lessons a method to produce time-altered video is also described. Completing this chapter, a summary of the uses and benefits reported by the scientific literature is presented.

6.1. Time-Scale Modification

When one hears about increasing the speed of an audio narration the first question is inevitably:

'Doesn't that sound like the chipmunks or Mickey Mouse?'

No. Time-scale modification or time-altered speech or compressed speech, has nothing to do with changing the playback speed of an audio recording. TSM is a signal processing method which effectively shortens a speech signal without modifying the tone of the spoken utterances. The result is a variation of the number of words spoken per minute.

Changing the WPM count of audio narration without changing the pitch has been around since the 50's. Garvey (1974) cited in Harrigan (1996) first started by 'splicing' slivers of magnetic audio recording tape and playing the reconnected shortened version. Initial electronic TSM apparatus were costly but today, readily available computer power makes it cheap to perform in real-time.

Steven King's "Lawnmower Man" (King 1978) strikingly illustrates the idea of a device allowing the learner to dynamically change the pace of audio-visual content delivery. Placed in an immersive digital environment, the learner controls the flow of images and related audio narration to phenomenal speeds. We are not implying knowledge can be so absorbed but maybe initial domain vocabulary and subsequent retrieval could be less time demanding. Indeed, "it has been shown the human brain can comprehend information that is delivered nearly twice as fast as the original speed of delivery" Carver (1982) cited in (Fulford 1992, p.1).

6.2. Pace of reading and speaking

Voice and silent reading rates are accepted (Junor 1992 and Harrigan 1995) to be respectively 125-175 and 275-300 words per minute. Speech is notoriously slow and contains many pauses. A speaker in a live lecture can be expected to think and pause repeatedly. Speech rate is often intentionally slows or accelerates to accommodate audience attention level.

We expect some gain in efficiency by providing the e-learner pace control over an audio/visual lesson through TSM. For example, Harrigan (1996) shows that once digitally recorded, speech can be time-altered and shortened by at least 10% without discernable difference. Figure 6.1 illustrates the range of WPM one can expect from normal speech, silent reading and time-altered speech.

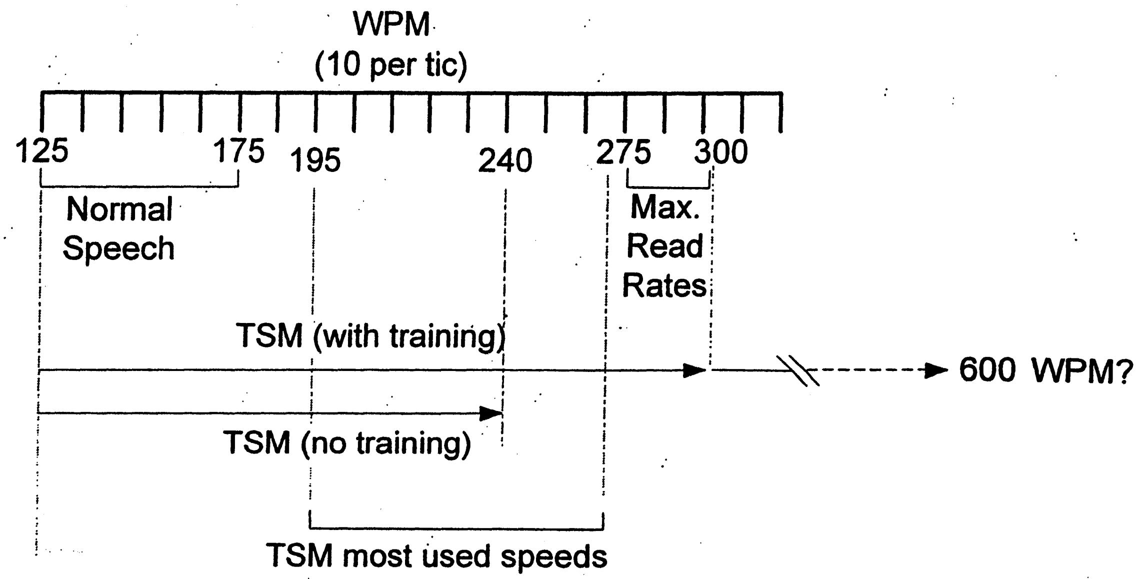

Figure 6.1. Range of WPM offered by written word, voice and time-altered speech.

The normal speech rate shown above is expressed as the "conversational speed (and) has been measured at 125-150 wpm. Professional voices have been identified as averaging 175 wpm" (Fulford 1992, p.2).

In a 2001 BYU deployment of audio-video TSM, most learners surveyed used speeds "between 1.3 to 1.8" (Galbraith and Spencer 2001) with the average rate at 1.6 or 240 WPM. This corresponds closely to "225 wpm (as) the rate reported to be preferred by learners" (Lars et Al. 1974) cited in (Fulford 1992, p.3).

A coasting compression rate has been observed when

> "the subjects started at a higher compression-rate based on their previous day's experience, and only made local adjustments over the duration of the session" (Omoigui, He, Gupta, Grudin and Sanocki 1999, p.140).

Learners can be expected to find their personally comfortable speed and vary sporadically according to content complexity.

One can see from Figure 6.1 that an audio/visual learner using TSM can match or even beat reading rates. The comprehension boundary shown at 300 WPM is provided by Carver (1993) and Foulke & Stiocht (1969): "there is an upper limit of approximately double the speed, at which point users reach their maximum rate of speech perception" cited in (Harrigan 1995).

Doubling normal speed is quantified by Foulke (1968) when he "proposed that compression threshold exists in the general area of 250 to 275 words per minute" (King and Behnke 1989, p.430). King and Behnke (1989) also quantify this upper limit of compression by stating that "a performance threshold was in effect at the 60% time-compression level" (p.439).

A 60% compression ratio represents the percentage of time saved by using accelerated speech through TSM and this unit is "the most commonly used" (Harrigan 1996, p.10) in literature. Thus if, as stated earlier, a normal speech rate is 150 WPM, 60% compression equates to 240 WPM. The following equation illustrates this operation:

$$150WPM * 1.6 = 240WPM$$

Even higher comprehension thresholds are stated in the literature. For example, "a listener can comprehend and process information at a rate up to or approximately 400 wpm" Silverstone (1974) cited in (Olson 1985, p.3). There is obviously a "continuing debate over the existence of a comprehension threshold" (King and Behnke 1989, p.430).

6.3. Limiting factors

The listening comprehension threshold is influenced by multiple factors. Indeed,

> "variables (such) as the speaker's sex, word production rate, voice pitch and timbre, and the signal to noise ratio of the recording all contribute to the limits of intelligibility" (Junor 1992, p.100).

We would add such factors as previous experience with TSM, depth of initial domain knowledge, mental abilities and motivation for example.

For instance, experiments have shown

> "extrinsic incentives during training for listening to compressed speech produces listening rates which were 2.2 times normal rate with no adverse effect on comprehension." Gade et Al. (1984) cited in (Gade and Mills 1989, p. 532).

This type of compression equates to an impressive listening rate of 330 WPM or, for example an approximate time saving of over 32 minutes when considering a 60 minute lesson. The following equations illustrate this efficiency gain:

$$60\,\text{min} * \frac{1}{2.2} = 27.27\,\text{min}$$ (Time altered lesson time)

$$60\,\text{min} - 27.27\,\text{min} = 32.73$$ (Time savings or efficiency gain)

Just as a physician can rapidly scan through a medical article, one can be expected to rapidly listen to a lesson concerning a familiar subject. Indeed, "the number of words processed per unit of time depends on the size of the chunk. This, in turn, is dependant on relevant prior knowledge in long-term memory" (Kozma 2001, p.155).

As well, familiarity with the speaking voice can also induce higher listening rates. For example, a user using

> "twice normal speed (…) was able to comprehend the speech at this fast rate even though she had no previous experience listening to time-compressed speech. This is most likely because the professor's speech was familiar to her and she (had) heard the material before" (Stifelman, Arons and Schmandt 2001, p.85).

Work on effects of pre-training with TSM listeners suggests the average listening rate could go past 300 WPM, up to a 600 WPM, with adequate training. Indeed, "TCS can be comprehended (between 500-600 wpm)" (Jones 1986, p.5).

6.4. Link between speech and reading rates

Literature hints at a correlation between the average silent reading rate and average listening rate users chose when TSM is available. Interestingly "the average rate of speech (200 wpm

± 25 wpm) corresponds closely with the average reading speed in the general population (200-250 wpm)" (Jones 1986, p.4).

Although there is some disagreement concerning precise WPM rates, the notion of a link between speaking rate and reading rate is accepted. Indeed, "the literature shows that people can listen at the same rate that they can read." (Harrigan 2000, p.78).

We wonder if the upper listening limit could be a by-product of physical limitations associated with the visual channel. Listeners may hit a 300 WPM threshold because the visual interpretation of printed characters has never allowed them to internally hear the narrative any faster. Removing the visual limitations, with adequate listening training, could one surpass this 300 WPM listening threshold?

6.5. Training with compressed-speech

The average user can acclimatize to TAS in a very short period of time (Gade and Mills 1989). For example, "Voor and Miller (1965) found that comprehension of compressed speech increased significantly over the first 8 to 10 min of listening" (Dupoux and Green 1997, p.915), and in general, "Mehler et al. (1993) showed that performance on compressed speech improve over time" (p.917).

During experiments, Dupoux and Green (1997) found that even limited exposure to TAS allows the user to adjust their ear. Indeed, "improvement in recall of compressed speech can occur with exposure to only 5 or 10 compressed sentences" (p.918) and in practice, "the higher the compression rate, the more exposure is required before performance reached a plateau" (p.925).

Gade and Mills (1989) would begin their tests by having all test users listen to a short compressed audio narration before starting the actual experiment. They found that "generally, the faster the practice passage, the faster the induced listening rate" (p.537).

Finally,

> "on the experimental side, there is the finding that, with practice, perceptual performance on highly compressed speech shows some measurable increase (Voor & Miller 1965)" (Dupoux and Green 1997, p.917).

We have presented the effect TSM has on the chosen rate of recorded speech. The next section will attempt to explain the speech signal manipulations performed by time-scale modification algorithms.

6.6. Speech signal treatment

Generally, time-scale modified speech sounds like the original except the words seem to be proportionally getting shorter as compression increases. TSM algorithms have evolved and diversified over the years. This section will explore the following time compression methods:

- Linear time compression (OLA, SOLA, PSOLA, WSOLA)
- Non-uniform or adaptive compression (Mach1)
- Pause removal
- Dichotic presentation.

These are compared and explained in He and Gupta (2001) and Harrigan (1996). Although practical impact on the quality of the time-altered signal will be presented, we will not delve into the technical implications of the speech signal treatment.

6.6.1. Linear time-compression

Linear time-compression entails speech signal manipulation is proportional from beginning to end. This type of TSM does not perform any linguistic analysis; it modifies a signal without considering what the utterances contain. The classic Fairbanks method will first be described followed by the more modern and popular OLA and SOLA algorithms. Finally, the signal manipulations involved to expand a speech signal will be presented.

6.6.1.1. Fairbanks method

The Fairbanks method (Fairbanks, Everitt and Jaeger 1954) involves removing thin slices of the signal and abutting the remaining segments together.

Figure 6.2 illustrates the effect of linear time altering using the Fairbanks method on a fictitious speech signal.

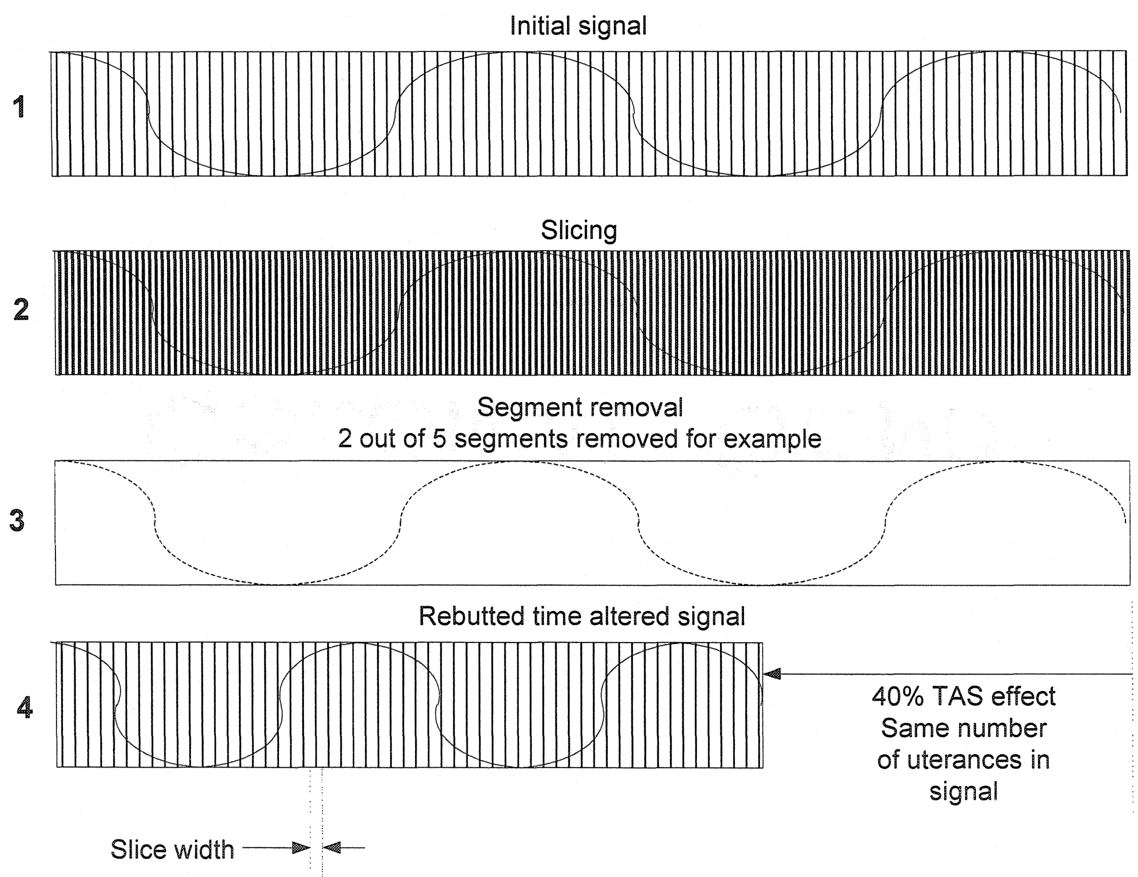

Figure 6.2. Fairbanks method time-altered effect on speech signal

The initial speech signal (1) is divided into fixed length segments of usually 30 to 100 milliseconds (2). In order to produce a 40% compression for example, 2 out of 5 segments are removed (3) and rebutted (4). Notice the number of utterances (peaks) and their amplitude is the same in the altered version. All words are present with the same pitch but each is felt to be uttered proportionally quicker and "the length of each chunk (or slice) is long enough to contain a complete pitch change but shorter than a phoneme" (Harrigan 1996, p.12). However, this simple method "produces discontinuities at the interval boundaries (…) audible clicks and other forms of signal distortion" (He and Gupta 2001, p.383).

6.6.1.2. Overlap Add (OLA) compression

A better method of joining speech segments together can be performed through algorithms based on a technique called Overlap Add (OLA) which improves the quality of the abutted signal weather or not signal slices are removed. Figure 6.3 visually presents an OLA type signal compression.

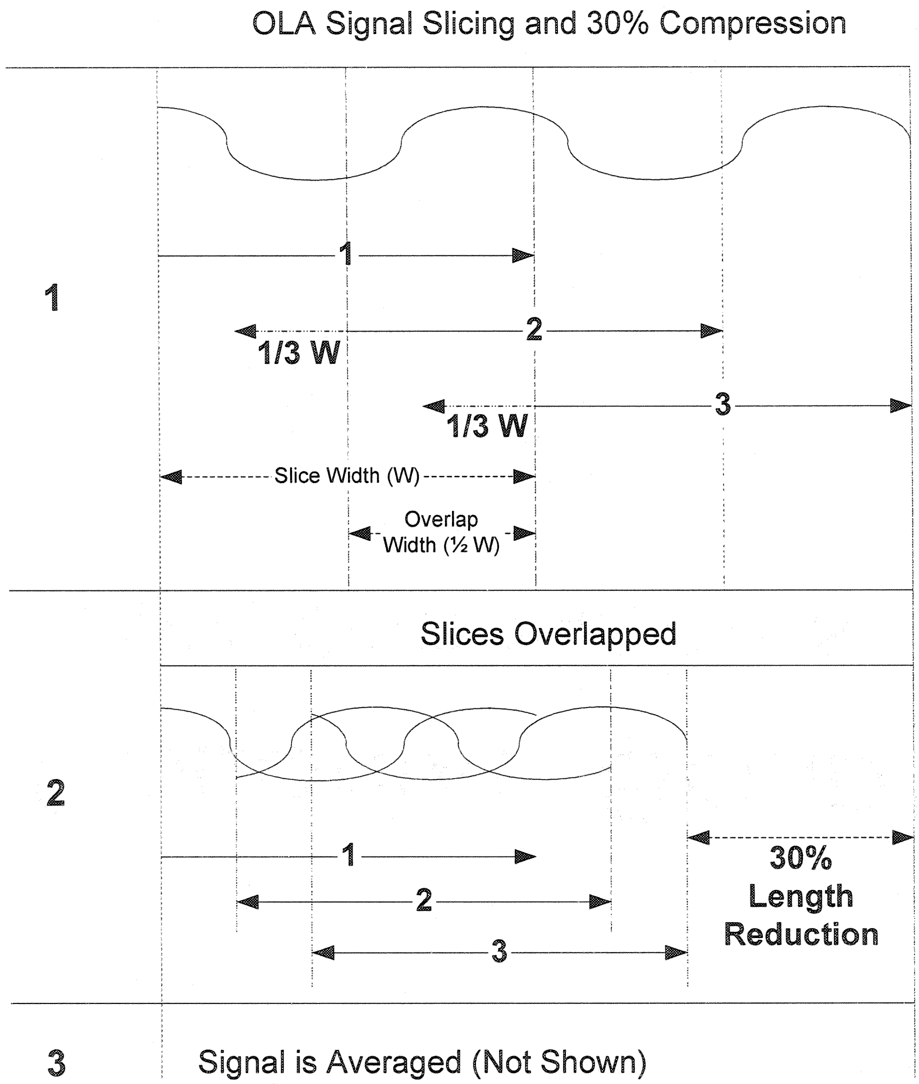

Figure 6.3. OLA signal compression

The signal is first divided into slices of length "W" with each slice overlapping the other by a fixed ratio (50% of W in our example). Each signal slice is then overlapped towards the beginning (left) of the signal by a fixed percent of its width (30% compression in this example) and finally, the signal is averaged. Unfortunately, this manipulation creates a signal with "low magnitude and extraneous pulses" (Harrigan 1996, p.15).

6.6.1.3. Synchronized Overlap Add (SOLA)

Further improving the compressed speech signal, the Synchronized Overlap Add or SOLA (Roucos and Wilgus 1985) time-altering algorithm produces an effect where "the beginning of a speech segment is overlapped and averaged with the previous speech segment at the position of the highest cross-section" (Harrigan 2000, p.79). Thus the overlap is no longer a fixed length but depends upon where the point of highest resemblance is situated between the two slice interfaces up to a chosen limit of evaluation. A detailed description of the SOLA algorithm and the related improvement of the audio quality is originally given by Roucos and Wilgus (1985) and explained in layman's terms in Harrigan (2000) and He and Gupta (2001).

The resulting increased speed of utterances does not create a "chipmunk" accelerated voice effect. Indeed "the compression process does not result in any changes in voice frequencies" (Paul 1992, p.72).

For the interested reader, the following are additional OLA based algorithms:

- PSOLA (Griffin and Lim 1984),
- WSOLA (Verhelst and Roelands 1993).

6.6.1.4. Linear time-expansion

Expanding a speech signal beyond its original length entails performing an OLA type manipulation like SOLA without removing slices and overlapping the signal segments towards the end (to the right) of the signal as shown in Figure 6.4.

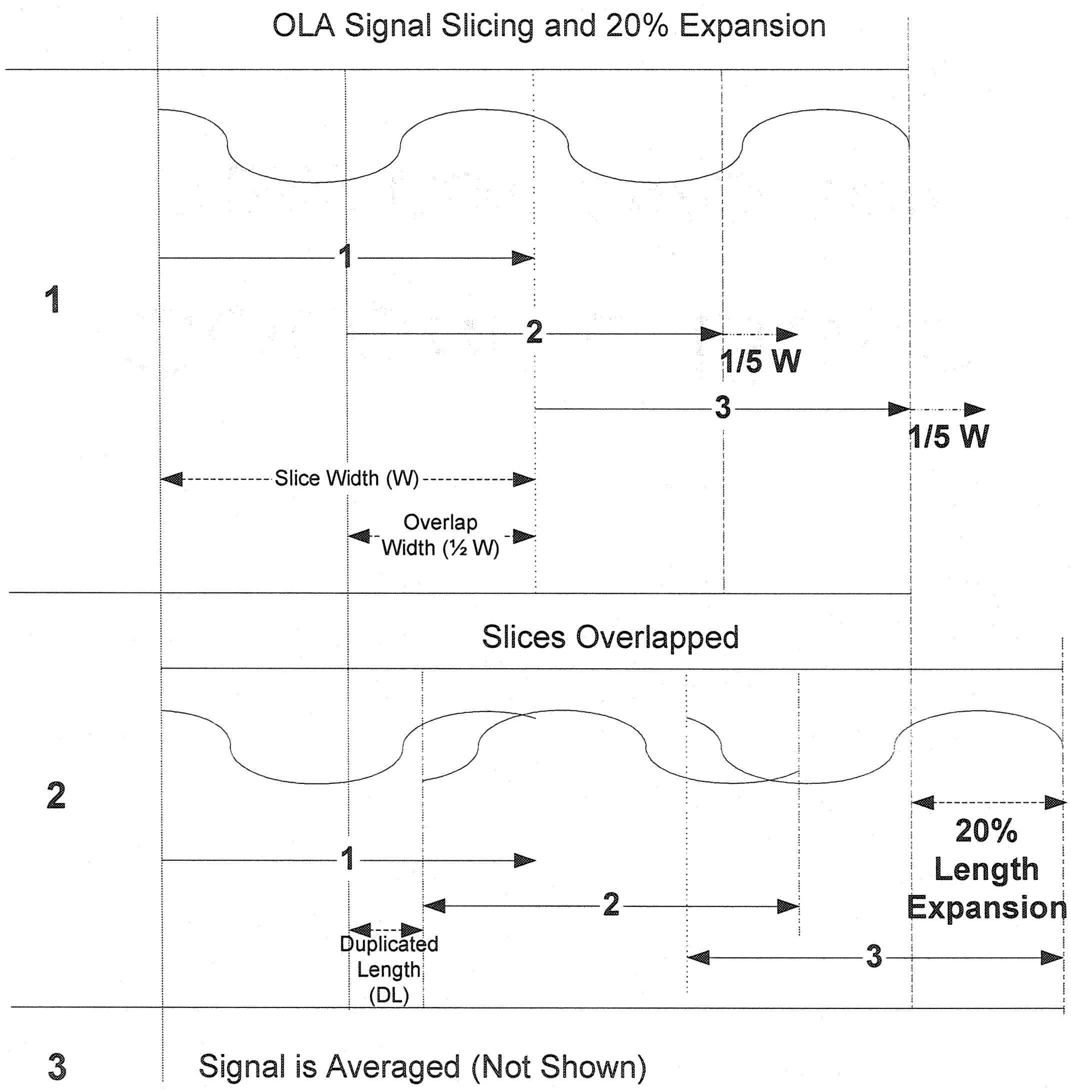

Figure 6.4. OLA signal expansion

In practice the resulting signal is stretched through periodic duplication and averaging of small signal segments. The duplicated length "DL" depends primarily on the expansion factor and the chosen slice width "W". For example, if one wished play a speech signal at 80% of its original speed or a 20% expansion, given usual values of W between 30 and 50 milliseconds, DL would be between 3 and 10 milliseconds.

The signal is shown to match original signal peak frequencies but there is a change in the amplitude "of the low frequency components of the time-expanded speech" (Dukes 1995, The signal is shown to match original signal peak frequencies but there is a change in the amplitude "of the low frequency components of the time-expanded speech" (Dukes 1995, p.19). Indeed, when using higher expansion rates, although intelligible, the speech acquires a noticeable change of pitch as compared with the original. In practice, expansion beyond 1.5 is not recommended and is simply difficult to follow.

6.6.2. Non-uniform or adaptive compression

Adding to the basic OLA compression, adaptive time-altering algorithms use linguistic characteristics of human speech to intelligently compress a speech signal. Examples of this type of compression can be found in He and Gupta (2001):

- MACH1 (Covell, Withgott and Slaney 1998)
- Variable TSM using transient information (Lee and Kim 1997).

Mach1, for example, uses the following linguistic heuristics to guide compression decisions (He and Gupta 2001, p.384):

- Pauses and silences are compressed the most
- Stressed vowels are compressed the least
- Unstressed vowels are compressed by an intermediate amount
- Consonants are compressed based on the stress level of the neighbouring vowels
- On average, consonants are compressed more than vowels.

The result is a variably compressed signal depending on the spoken utterances. Since they account for the linguistic characteristics of the voice message, non-uniform algorithms do provide greater intelligibility but demand more processing power.

After comparative experimentation with users, He and Gupta (2001) concluded "the benefits of the non-uniform algorithms are more pronounced at a higher compression rate" (p.388). The authors observed that most users generally do not use compression rates where these benefits would be apparent and thus advise the use of classic linear compression. The subjects used in the study had no prior experience with TSM. We wonder if higher compression rates chosen by users with prior domain knowledge and extensive training with TSM could benefit from these non-uniform algorithms.

6.6.3. Pause manipulation

To compress a speech signal further, one can also remove or shorten pauses which are beyond a certain length of time. For example, He and Gupta (2001) showed pauses can be gradually cut to shorten the speech signal by 15-20% without any signal distortion to the individual words. In support, "studies have shown that the audio with the pauses deleted can be understood as well as or better that regular speech" (Harrigan 1995, p.363).

Pauses do have cognitive use but, they

> "are only useful when they occur between clauses within sentences, pauses within clauses are disrupting. Pauses suggest the boundaries of material to be analysed, and provide vital cognitive processing time" (Arons 1997, p.4) and "brief pauses between sentences making up the discourse was facilitating, especially at a high rate" (Paul 1992, p.72).

A general pause removal guideline is provided in Harrigan (1996) by Minifie (1974) when he states "pauses over 500 milliseconds (ms) can be shortened to 500 ms without loss of comprehension" (p.11).

6.6.4. Dichotic presentation

When ear phones are available one can provide the listener with a dichotic presentation. This technique involves "playing the standard sampled signal to one ear and the 'discarded' material to the other ear" Scott (1967) cited in (Arons 1997, p.4).

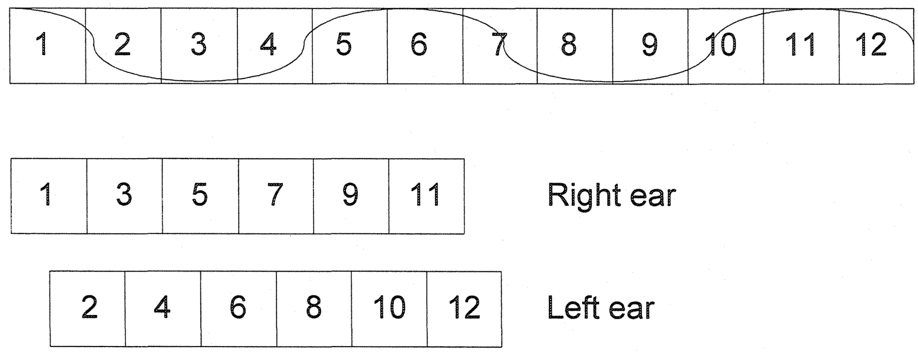

Figure 6.5. Dichotic audio presentation

The initial signal is again divided into fixed-length segments (12 in the example above). The segments are then split up sequentially and sent in alternate order to each ear.

Dichotic presentation "takes advantage of the auditory system's ability to integrate information from both ears. Intelligibility and comprehension increase" Gerber (1977) cited in (Arons 1997, p.4). We believe this type of compression shows promise and have not found experimental applications in education.

We have presented the different methods available to time-alter an audio speech recording. For audio/visual educational material, the related visual content can be static (photographs, illustrations, graphs, etc.) or dynamic through video or computer generated animation. Compression of static visual content in synchronicity with the audio narrative simply means shortening the time-exposure but, video compression is another matter. The following section presents a summary of research concerning time-altered video.

6.7. Time-altered video

In his 1996 doctoral thesis presented at the department of education at the University of Toronto, Kevin Harrigan states "the use of time-compressed video has not been studied extensively by researchers" (Harrigan 1996, p.19).

He describes a method to produce compressed audio-video content. First, the audio and visual tracks are digitally separated. Using the SOLA algorithm, the audio is time compressed

6.7. Time-altered video

In his 1996 doctoral thesis presented at the department of education at the University of Toronto, Kevin Harrigan states "the use of time-compressed video has not been studied extensively by researchers" (Harrigan 1996, p.19).

He describes a method to produce compressed audio-video content. First, the audio and visual tracks are digitally separated. Using the SOLA algorithm, the audio is time compressed to a desired level of 25% for example. The video track is then compressed by systematically removing the same percentage of frames. In this case, 1 out of every 4 frames would be removed. After re-integration of the audio and visual tracks, the resulting audio-video is 75% of its original length.

Video compression using a linear algorithm like SOLA does not have a direct effect on the number of frames contained in the video. Each frame is proportionally shown for less time thus the movements are naturally fluid yet accelerated. On the other hand, pause manipulation or non-uniform TSM may introduce visual discontinuities. For example, since people often listen by visually "reading" a speaker's lips, the removal of video frames caused by pause reduction or non-linear TSM diminishes fluidity and interpretation of this visual cue. Finally, pause removal in video creates additional breaks in transitions. By removing a 500 ms pause for example, an actors' arm can surprisingly appear in a different position in space.

The compression capabilities of audio/video are less than strictly audio presentation. Indeed, "it appears that speech alone can be sped up more than video can" (Harrigan 1996, p.19). He goes on to quantify this observation stating "video can be time-compressed more than 33.3% but less that 50% without loss of compression" (Harrigan 2000, p.79).

Recent research does show promise for audio-video compression. Indeed, "people can comfortably view videos at speeds 1.5 to 2 times normal (speed)" (He, Grudin and Gupta 2000, p.133).

6.8. Usage and benefits

Over the years, time-scale modification has received much attention from a wide range of authors. This section presents a sample of the literature concerning the use of time-compression in education.

6.8.1. Performance and efficiency

One can expect a certain impact on learning performance and efficiency. "The use of TCS appears to aid in focusing concentration and to require less effort to enhance performance" (Jones 1986, p.6). As well,

> "regardless of age, intelligence, or sex, the time-compression of recorded audio information has been shown to increase the performance of students" Myers (1978) cited in (Junor 1992, p.101).

Simply put, "time compressed speech can make listening to recorded speech considerably more efficient" (Paul 1992, p.72).

Additionally, in an experimental study performed at Microsoft Research, it has been shown pause removal and time-altered speech "increased the fraction of content they (users) watched once or more from 33.1% to 48.2% (resulting in an) effective playback speed of 137.1%" (Li, Gupta, Sanocki, He and Rui 2000, p.173). This is equivalent to a 37.1% compression rate or time saving. The audio/visual contents watched by the test subjects were comprised, among others, of classroom lectures, conference presentations and news.

TSM benefits can be increased by considering certain factors during audio script design and production. "The use of longer words in the text has also shown to be an aid in the intelligibility of the compressed passages" Heise (1971) cited in (Olson and Berry 1982, p.8). Additionally, "articulation rate contributed most to the listener's concept of speaking rate" Lane (1976) cited in (Olson and Berry 1982, p.13) and "material of moderate difficulty should be used rather than very complex material when speech is to be compressed" Watts (1970) cited in (Olson and Berry 1982, p.9).

Comparing printed text to the equivalent audio narrated material, experiments have shown "the comprehension of auded material has been superior to that of identical printed material up to about 200 wpm" (Junor 1992, p.97). Taking a normal speaking rate of 150, 200 WPM represents speech approximately 1.3 times normal rate without loss of comprehension. This represents a 30% efficiency gain for a learning media said to be superior to printed text.

Olson (1985) states TSM usage is highly recommendable for the efficiency of learning tasks involving what he calls spatial restructuring, spatial analysis and comprehension. A caveat is expressed concerning recall of specific technical information. He states "normal rates are superior to compressed rates when information is technical in nature" (p.12). In this case, normal rates could indeed be used for the first time but one can assume subsequent review sessions would be sped up to some extent. Additionally, this only concerns verbatim recall of technical terms.

Even when no visual content is present, "text augmented with compressed speech is still more efficient" (Fulford 1992, p.15) and "the faster material is presented the more apparent it becomes that simultaneous reading-listening is the optimal mode" (Jones 1986, p.5).

6.8.2. User preference

Once dynamic pace control of content delivery is offered to learners they voice a strong preference for its ongoing availability. For example, Rippy (1975) found a

> "strong preference for compressed speech among college students. (He assumes) students may prefer compression if they were permitted to control and vary the rate" cited in (Olson and Berry 1982, p.17).

He concludes these advantages make "time-compressed speech a signification compliment to 'speed reading'" (p.27).

As well, Omoigui et Al. (1999) found that "across all five videos, the savings in task-time was 22%. (…) the subjects liked the time-compression feature very much" (Omoigui, He, Gupta, Grudin and Sanocki 1999, p.143), and "Gould and Boise (1984) have shown that not only do learners comprehend the time-compressed speech, but they also prefer it to speech at normal speed" (Harrigan 1996, p.17).

More recently, a 2001 TSM deployment for an introductory accounting course at Brigham Young University showed some very promising results among the 650 survey respondents. The learner feedback evaluations showed

> "many students reported that accelerated viewing rates kept their attention longer, and provided significant time savings. Most users preferred an accelerated speed of at least 1.5 times normal, and several users reported watching the lecture material comfortably at speeds of twice normal and higher" (Galbraith and Spencer 2001, p.2).

The students also used TSM as a searching and scanning tool to "skim through a video presentation without 'skipping over material' (p.4).

Overall, 94% of the surveyed BYU students rated the usefulness of TSM between values of 5 to 7, a value of 1 indicating 'not useful' and 7 representing 'very useful'. Students reported being "less distracted by their surroundings, and more interested in the lesson material" (p.8).

The report authors note that "the ability to continually (or intermittently) decelerate or accelerate lectures according to individual comprehension level and interest fundamentally changes the nature of video from a traditionally 'lean back' passive medium, to a "lean forward" active delivery medium" (p.9).

We believe this 'active' learning benefit to be rather significant. No matter how much one invests in simulation and interactive practice problems, a student requires an initial domain vocabulary acquirement stage. This usually involves low interactivity since learners are asked to accept a base of new terms and definitions before one can begin practical applications. As the BYU experience shows, dynamic pace control renders such passive ingestion more interactive and accentuates learner control over the learning process. As is generally accepted

> "the use of software, which provides high levels of user control will help students feel that they are instrumental in determining the pattern and process of the learning experience, i.e. in developing a sense of ownership" (Squires and Preece 1999, p.478).

6.8.3. Intelligibility, comprehension and retention

Comprehension and intelligibility are different. One can understand words without comprehending implications to prior knowledge. Once a speech signal is intelligible, learning occurs when one comprehends the meaning and retains key elements in long term memory.

Retention refers to verbatim memorizing of content. Experiments have demonstrated "retention of speeded information is as good as retention of information presented at normal speed" (Olson and Berry 1982, p.9).

Where integration with known facts and concepts is required King and Behnke (1989) cite studies where, as one might expect, it is clear there is "a progressive decline in comprehensive listening effectiveness with increases in compression" (p.437). They quantify "60% compression is the level at which comprehension is sharply curtailed" (p.430). This rate is the most conservative we have found in the scientific literature. As we have seen in section 6.2, some question the very existence of a comprehension threshold.

For the sake of argument, we will accept this 60% compression level as the upper compression limit to comprehension. This represents an increase in WPM rate from the normal 150 to a time-altered 240 or, a 36 minute savings for a 60 minute lesson. We can see even the most conservative estimates show TSM can cut lesson time by more than half. Given the choice, we would be curious to hear students explain why they would rather not save the time.

Adversely, concerning intelligibility, "as long as the speech is not speeded up by more that 50% the listener will be unaware that there has been an electronic alteration of the original recording" Labarberra and MacLachlan (1979) cited in (Fulford 1992, p.3). This rate seems high in light of our own experience with TSM. We cite this passage in order to reaffirm a certain efficiency gain is possible without users noticing the speech signal has been altered. We would venture to quantify this *unnoticeable* rate of compression at no more than 30% but further experiments would be required to shed light onto this rather subjective question.

As the compression rate increases, the voice signal becomes proportionally difficult to understand. This problem can be alleviated through multi-modality since

> "it is evident from previous research that the loss of comprehension due to signal distortion and to the limited time for processing may be compensated by the addition of pictures redundant to the discourse" (Paul 1992, p.75).

After experimenting with comprehension of different material densities, Paul (1992) concludes "that comprehension of compressed speech is not a function of number of propositions making up the discourse" (p.76). In other words, according to Paul, even a dense and complex narrative can just as easily be understood as a simple children's story for example.

Indeed, content judged technical in nature can be safely understood by one unused to TSM.

> "Shields (1975) (…) demonstrated that listeners untrained in listening to compressed speech could accurately identify the subject matter of highly compressed technical communication compressed to 1.5 times the normal rate of 150 words per minute (wpm)" (Gade and Mills 1989, p.531).

In practice, this would mean a 30 minutes time saving for an hour long audio lesson.

6.8.4. Usage

In his 1996 doctoral thesis, Kevin Harrigan established the just noticeable difference (JND) of speech compression at "approximately 5% (compression) as this is the minimum JND in all the conditions" (Harrigan 1996, p.88).

Alternatively, "psychological factors may be primary driving forces to push for the lower-latency continuous granularity functionality" (Omoigui, He, Gupta, Grudin and Sanocki 1999, p.141). In other words, these authors suggest users would prefer immediate impact on speed of delivery through a fine granularity time-scale modification widget. Practically, allowing users to control compression through a slider would be preferred over a set of radio buttons with predefined rates thus giving the learner real-time, precise control over the speed of delivery.

Although this research does not delve into the potential benefits of decelerated speech, Junor (1992) observed that "aged subjects (70 years and older) benefited from deceleration of the passages to 80% of normal speed" (p.101). Uses for decelerated speech are also expressed by Barry Arons. For the past 10 years, he has concentrated his efforts on audio skimming and navigation and states "it has been informally found that playing the segments at normal speed (…) even slowing down the speech, is useful when skimming unfamiliar material." (Arons 1997, p.10).

Interestingly enough,

> "it seems that mature age (around 28-30 years) students are more capable listeners at any presentation rate than the typical undergraduate student. As the age composition of distance education students tends towards mature students, they are likely to be able to realize on the potential of time-compression of recorded information." (Junor 1992, p.102).

Much discussion can be generated from this last statement. Junor's article is cited in numerous later publications but none seem to comment this assertion. It entails listening abilities generally improve well into the late twenties. This would be interesting to verify.

Demonstrating the review potential of compressed speech, students viewing a previously attended lesson used "the fastest speed (…) 75% of the time." (Harrigan 1995, p.367) and, "a viewer who joins late could review the beginning of a talk at a compressed speed and 'catch up'" (He, Grudin and Gupta 2000, p.134).

Time altered speech should by no means be seen as a magic pill to learn better and faster. This technology should "be employed in concert with other features to give users the power to quickly interact with multimedia content" (Omoigui, He, Gupta, Grudin and Sanocki 1999, p.142). These could include summaries produced by the domain expert and audio/video navigation capabilities for example. The SpeechSkimmer application (Arons 1997) and digital video browsing software (Li, Gupta, Sanocki, He and Rui 2000) present some of these features.

Generally, it is said that "time compression had few negative effects." (Schwab and DeGroot 1993, p.82) and "given all the positive aspects of time-compressed digital-video it would be reasonable in the future to have time-compression controls built into the user interface widgets for access to all digital video" (Harrigan 1996, p.94).

This chapter presented TSM as a means to provide user control of e-content pace of delivery. As we have seen, much positive endorsement has been given by a significant number of scientific authors. Although we are inclined to believe this technology can do no harm and should help the online student, some questions have arisen and are presented in the following chapter.

CHAPTER 7: RESEARCH QUESTIONS

Cognitively efficient e-content should generally be transmitted using an audio narrative. This being said, could the use TSM render audio/visual information consumption more efficient and reduce the time of new concept acquirement and subsequent review? Would TSM be appreciated by the e-learner? In light of the scientific literature review presented in chapters 5 and 6, we are inclined to believe so.

We now present two research questions emanating from the previous literature review. First, the use of TSM with specifically designed online tutorials of a technical nature has not received empirical attention. Second, a more elusive question is expressed concerning the current lack of availability of this seemingly beneficial educative technology.

7.1. TSM with a Technical Tutorial

We wondered what the effect of TSM would be in the case of an audio/visual tutorial as opposed to pre-recorded lectures. As well, what use would TSM have with technical material delivered through a high visual density tutorial? Content delivery speed might even be slowed below normal rate in this case. This would decrease time savings but allow the e-learner to stave off cognitive overload and possibly not have to repeat the same lesson passage as many times as when using normal speed.

In comparison, recorded lectures generally offer low visual information density. Aside from the speaker's lip movements and facial expressions, overheads might be used and formulas shown or written on the board. Overheads are often used as lecturing cues for the speaker rather than stand alone content. Text on a board is inherently written slowly and progressively. A specifically designed tutorial on the other hand would not contain these inherent pauses and content delivery rate changes.

To answer this questions we offer a novel empirical study adding to the general scientific support for user pace control of e-learning content delivery and hope it may convince a few more people of its benefits.

7.2. Lack of Mass Availability

Given the strong case for TSM use in education, one wonders why this functionality is not readily available in current e-learning offering. A massive amount of experimental literature touts the benefits to learning and recent empirical studies strongly suggest efficiency gains and user preference. The technology is stable and cheaply usable with any personal computer. This functionality, it would seem, can only offer potential benefits without negative side effects. If an individual e-learner does not want to use the functionality, he or she doesn't have to use it and can choose to keep normal speech rate. So why is TSM not offered with online training material? The answer to this question has proved to be somewhat elusive and we will present some tentative explanations in section 7.4.

7.3. Empirical experiment

This section describes the use of TSM by a pilot group of students enrolled in the engineered materials class at l'École Polytechnique of Montreal. We begin by exposing the different ways to measure TSM induced learning efficiency gains and the context of the empirical study is described followed by a description of the actual event. Finally results and findings are presented.

7.3.1. Measuring efficiency gains

In order to measure the potential time savings generated by dynamic pace control, we would theoretically compare the normal time a student requires to watch and listen to a multimedia lesson with the time used when pace control is available. This is an impossible task since we will never know the 'normal' time once we provide the learner with pace control. You can only be a novice once.

The alternative is to calculate average compression rate a student chose and compare with the fixed normal rate (usually approximately 150 WPM). This compression ratio provides an idea of the time savings compared with the same lesson watched at the normal rate. The following equation produces the average compression rate used by the learner.

$$\overline{CompressionRate} = \frac{\sum(UserTime(i) * CompressionRate)}{\sum UserTime(i)}$$

(Omoigui, He, Gupta, Grudin and Sanocki 1999, p.139)

This type of calculation requires usage of the TSM control to be continuously logged. Each compression ratio chosen must have a time stamp when selected. The difference between a time stamp and the next gives the UserTime(i) value for that compression rate.

Continuous logging presents practical problems. For example, no one can guaranty the user is actually watching the lesson at any point in time. The learner can receive a phone call or be interrupted and will not necessarily stop the lesson. Thus, the above equation can only be used in a controlled experimental setting.

For empirical study purposes we rely on user surveys. As in the Brigham Young University TSM deployment (Galbraith and Spencer 2001), participants are simply asked which compression ratio they used most of the time. This is relatively reliable since there are indications (Omoigui, He, Gupta, Grudin and Sanocki 1999) users find their coasting speed after a short adaptation period. We recognize this will produce be a rough estimate of the time savings associated with TSM use.

7.3.2. Empirical study context

The engineered materials course is at the forefront of computer supported teaching tools at l'École Polytechnique of Montreal. For the past 3 years, students have access to a complete CD-ROM based tutorial. Surveys have shown this multi-media reference is used and appreciated by 94% of roughly 1000 students per year who take this mandatory, 2 credit first year course.

The contents are delivered primarily via a professionally recorded audio narration at an average rate of 150 WPM with supporting computer generated animations and videos. Mathematical formulas also appear in accordance with the narrated explanation. Additionally, the online learner has access to interactive 3D models, diagrams and still photos and a bank of practice exercises.

The tutorial's first chapter describes material characterisation methods and contains 21 distinct screens or pages. Each page offers approximately 1 to 3 minutes of continuous audio narration synchronized with computer generated animations, video or static graphs and, mathematical formulas.

Interactivity takes the form of standard video type controls (play, pause, jump to beginning and end) and ability to jump to the previous or next pages. Two pages offer a link to online exercises and, three provide a link to still photograph examples. Finally, one page offers no audio narration where the user is expected to read a short paragraph of text and interpret the associated graph.

The minimum time required to view and listen to the complete chapter is approximately 30 minutes. Adding some time to look at the three photo examples and interpret the strictly visual page, we would expect about 35 minutes to complete this lesson without performing the exercises. These can be done through in a separate interface or paper support.

This online tutorial was originally developed using "Asymetrics" now called "click2learn.com". This software is designed to assemble online training presentations based on a time-line metaphor allowing synchronization of audio and visual elements. This software does not offer any standard TSM capabilities.

This empirical study differs from those found in the literature. Notably, the recent Microsoft (He, Grudin and Gupta 2000) and BYU (Galbraith and Spencer 2001) deployments used TSM along with digitally recorded lectures. We can wonder if the scripted nature of a tutorial reduces inherent inefficiencies related to live lecturing in a classroom setting. This should result in a lesser TSM efficiency gain for our tutorial.

Additionally, the tutorial content is judged technical in nature and contains section with high visual information density along with audio narration. In some sections, the learner is asked to listen to the narration, interpret a 3D animation and understand a mathematical demonstration of the phenomenon relating to unfamiliar concepts such as constraints and induced elastic and plastic matter deformations. Although Olsen (1985) supports efficiency gains using TSM even for technical content, we wonder if students might not actually slow the pace of delivery below normal for visually dense segments.

7.3.3. Offering TSM to a pilot group

In order to verify if TSM is a useful feature for the learner, we offered pace control for the first chapter of this computer tutorial and measured impact on learning through a survey. Providing dynamic compression capabilities using the current software would force us to pre-select a set of compression ratios which would be available to the user via a set of radio buttons. The associated pre-compressed audio and video files would then be synchronously played.

This goes against the notion of complete user control of TSM granularity and minimal lag when switching compression ratio. Additionally, the fact that this would not be real-time audio/video compression demands greater storage requirements. For each compression ratio we wish to offer, the associated pre-compressed audio and video files must be available on the provided CD-ROM.

A better, cheaper alternative is to convert the complete chapter 1 lesson into digital video format. This would create a single audio/video file which could be viewed using Windows Media Player™ or RealPlayer™. These multimedia viewing applications provide real-time audio/visual TSM through a third party plug-in named 2xAV from Enounce Inc. This plug-in allows single percent TSM acceleration and deceleration through a slider interface widget. As well, these multimedia players offer standard video controls like the current tutorial.

The drawbacks of this second alternative are as follows. First, the student is required to install and use one of the multi-media players (both are free downloads for educational purposes) and the associated plug-in (free 7 day trial). Slightly more troubling is the fact that clickable links within the pages will not be active. Practically this entails that the two links to exercises and the three links to still photograph examples will not work.

In the case of the exercises, we believe this can be alleviated by warning students and advising them to access the exercises through the usual interface or paper support. As for the still photograph examples we propose to simply show them for a fixed duration. For example, they would appear for 10 seconds keeping in mind the learner can simply accelerate or decelerate this time of appearance in accordance with his or her understanding the photo content.

This version of the first tutorial chapter is offered in addition to the standard tutorial. The pilot group was instructed in TSM usage and associated drawbacks of the current test lesson. They covered the material on their own time and completed the survey shown in appendix 1. Additionally, in order to collect qualitative data and anecdotes, we asked for student volunteers to be observed while using TSM

7.3.3.1. Live observations of usage

In order to gain realistic insight into TSM usage, we observed a subset of 5 students while they studied the online tutorial as part of their overall class preparations. In this type of observation "the critical data furnished by the experience is not performance level as such but the methods by which the performance is achieved" (Vygotsky 1978, p.13). These students agreed to set an appointment and watch the chapter tutorial in my presence. This group was comprised of two females and 3 males ranging between 19 and 26 years of age. One was a PHD level student while the others were first or second year undergraduate level.

To allow each subject to become comfortable with the observer's presence, we engaged in light conversation before each observation session began. Usage of the Enounce plug-in and WMP was demonstrated and practiced by the user while I left the room for about 2-3 minutes. Generally, I acted as a silent observer although some questions were asked to understand reasoning behind specific actions taken by the student. Hand written notes were taken throughout the process. I usually asked several questions after the tutorial was completed and solicited the student's impression and comments.

Generally, all subjects demonstrated very different usage behaviours with the technology but, all tended to use delivery speeds other than normal. All had attended the first class which covered some of the tutorial material. This led most to accelerate through passages felt to be already understood. For instance, one student used speeds up to 1.6 times normal considering the material to be of questionable relevance but, generally the top speed used was 1.2 which would support the assertion TSM increases learning efficiency.

Due to the high visual information density, one particularly assiduous PHD student slowed the speed down to 0.8 times normal. He explained he ordinarily would have played back these complex passages many times until understood. This behaviour was also observed with three other subjects. They also slowed the delivery speed when visual density was highest.

The number of speed changes was also highly variable from one subject to the other. One set the speed to 1.1 times normal and never changed it again. She was observed being uninterested through some passages but felt she would rather passively wait for the next complex segment. All other subjects dynamically modified the speed (5 to 15 modifications) based on visual information density and comprehension. For instance, when the audio narration was strictly accompanied by a video presentation, speed was generally increased only to be reduced when a mathematical formula demonstration and related 3D model were explained by the narrator.

During post observation conversations, all expressed enthusiasm for the technology. They felt its value was not necessarily associated with time savings but with the ability to match speed of delivery with the dynamic complexity of the contents.

After observing these five subjects it is clear one can not pretend to have a predictive model of student behaviour with online tutorials. Some look for a passive audio/visual experience while others wish to interact constantly with the contents and software. Therefore, these observations reaffirm the value of high user control over content presentation and delivery. Without a doubt, TSM supports this objective.

7.3.3.2. *Survey results*

We collected filled surveys forms over a period of three weeks following the initial test chapter distribution and received a total of thirty (30) from the ninety (90) initially distributed. This low return rate of 33% can be explained in part by the low overall student attendance level after the initial class. Indeed, both professors estimate actual attendance at 50% compared with those who attended the first introductory class. Some students may have simply dropped this fast pace and dense summer course while others may have chosen to rely strictly on the handbook and CD-ROM tutorial. We are told this phenomenon is historically not surprising for this course.

We statistically analysed the collected surveys and Table 7.1 presents the results to questions 1, 2 and 4. Complete results can be found at appendix 2.

Table 7.1. Summarized survey results for questions 1, 2 and 4

Survey question	**1. Usefulness (Scale of 1 to 7)**	**2. Most used speed**	**4. Content difficulty (Scale of 1 to 7)**
Average	5.14	1.28	4.5
Standard Deviation	1.3	0.2	1.25
% Std Deviation	25%	15%	28%
Low	3	0.9	1
High	7	1.7	6

Overall, as shown by question 4, lesson difficulty was perceived to be of moderate level compared with other subjects and past experiences at l'École Polytechnique. This rating lets us assume the contents are of average difficulty and adequately represent the students usual study tasks.

Concerning perceived usefulness, the first question results indicate the functionality was generally thought to be beneficial to the learning process and indeed, 90% of survey respondents assigned a rating of 4 or more. Some students felt TSM should be offered with all twelve (12) course chapters and others commented that in general, features adding user control over the software are beneficial. A few students had simply not modified the listening speed and stated TSM was not very useful. As expected and observed, some looked for a passive multimedia lesson but most interacted with the tutorial interface including TSM.

The average speed reported most used was 1.28 which rounded to 1.3 times normal rate. In light of its low relative deviation of 15% we feel this reported value is statistically significant and translates to an approximate five (5) minutes time savings for this half hour lesson. This efficiency gain reported by a majority of respondents would be considered a significant endorsement for TSM in this context.

Question three (3) allowed the respondent to select multiple answers and Table 7.2 presents the results which provide nuances against hasty conclusions pertaining to the previous positive results.

Table 7.2. Summarized survey results for multi-select question 3

Multi-select Q3	**a) Kept same speed**	**b) Changed according to difficulty**	**c) Used slow speeds**	**d) Found my speed at start**	**e) Re-watched a segment**	**f)Used TSM to search**
N	5	21	10	6	11 (2-3 times)	9
% of group	17%	70%	33%	20%	37%	30%

Care must taken not to promote a 30% efficiency gain too strongly since the vast majority of students report varying the speed and not consistently keeping the same pace. As option "b" of Table 7.2 shows, 70% of respondents state they dynamically modified the speed according to the content difficulty. As was observed during the live observations, some specifically commented they increased the pace during segments where only a video was presented and returned at or close to normal during segments containing mathematical demonstrations. Thus, we can assume the actual average pace of instruction might not be adequately represented by the speeds reported in question 2.

Additionally, as options "a" shows, only 17% stated they kept the same pace throughout the lesson and few selected option "d". This suggests most varied the pace of instruction during the lesson. These findings seem to contradict the results found at Microsoft Research by Omoigui (1999) where they reported most users kept the same compression rate after an initial period of adaptation. Since our study only pertains to a half hour lesson it may be that a longer usage period, such as the 2 day experiment conducted at Microsoft Research, would produce the same behaviour.

A third of the surveys reported speeds lower than normal were used for at least one segment. This is shown in option "c" of Table 7.2 and would necessarily indicate that for some segments, normal speech rate of 150 WPM would be considered too fast for a significant portion of students. This is corroborated by selections made to option "e" where 37% state they watched the same passage 2 or 3 times. This would indicate TSM usefulness even without a measurable efficiency gain since we can expect the same passage to be watched less often if presented at a slower pace.

Finally, as option "f" shows, 30% of the respondents report using TSM to find a particular passage. This type of use may be more significant prior to an examination but our study time frame did not include such an event. This tutorial chapter was watched by most subjects after the contents were presented by the professor during the first live lecture. We assume some students accelerated the tutorial lesson to find passages they had not understood through teacher explanations.

Although voluntary survey responses must not be taken as a precise representation of reality we believe it is safe to assume students generally felt TSM was useful and dynamically modified the pace of delivery throughout the lesson. Although an approximate 30% gain in efficiency was found we feel it is sound to stress the positive impact of the increased user control of the learning experience since actual time savings may not be represented by reported average speeds.

7.3.3.3. Effects of Compression/Expansion on Cognitive Load

We were surprised to find expanded speech was used by observed students and reported in the surveys. As mentioned in section 6.8.4, Arons (1997) stated expanded speech can be useful when viewing unfamiliar material and Junor (1992) reported the elderly benefited from a rate reduction of 80% of normal or 120 WPM. None of our study subjects were elderly and all had previously attended the lectured class concerning the material before reviewing the CD-ROM lesson. In theory, the material should have been "familiar" to them and not necessitate speech expansion.

Although not measured, observations point to chosen listening speeds being inversely proportional to visual information density. Let us state the total cognitive load (CL) demanded by the audio/visual lesson is induced by the audio narration (AL) and the visual information loads (VL). Assuming AL is proportional to the WPM count and constant, only the cognitive load imposed by the visual channel forces CL to vary. The following equation illustrates this relationship.

$$CL = AL(Const) + VL$$

Each independent sub-section of the half lesson contains a variable amount of visual information the student is asked to interpret and integrate with the continuous audio narration. For example, some sections show a video, others show a 3D animation and a mathematical demonstration, others a dynamic graph, a related video of the phenomena and written definitions. Establishing a way to evaluate transient visual density of a dynamic online lesson does not fall within the scope of this research but generally, the more you show, the more you force the student to slow down or rewind.

It is interesting to note that visual information density is highly variable throughout the half hour chapter but the imposed speed of audio content delivery is a constant 150 WPM. We hypothesize that in a normal setting the tutorial student is overwhelmed by some segments and inadequately stimulated by others. The graph shown in Figure 7.1 is based on fictitious data in order to illustrate this phenomenon.

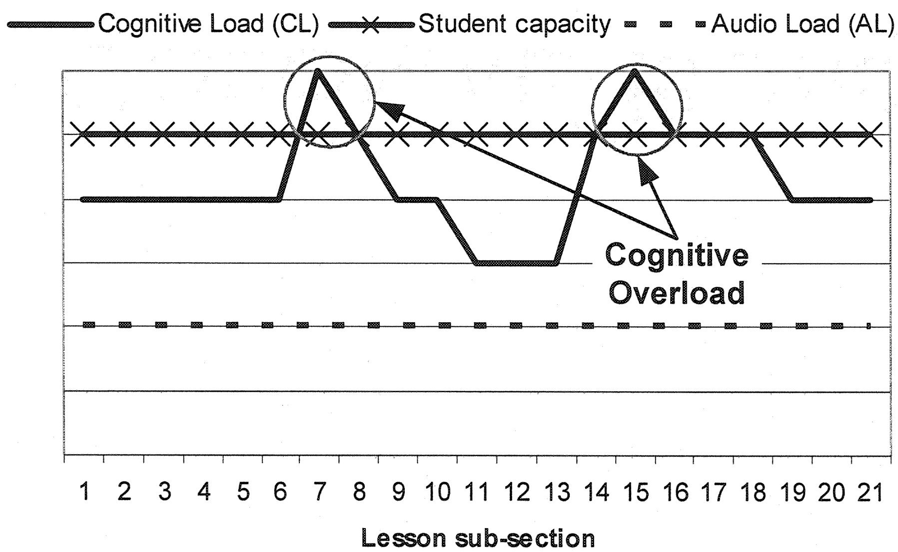

Figure 7.1. Cognitive load induced by audio-visual lesson

In this example the cognitive load induced by the audio narration is represented by the dashed line and constant. The visual information density is variable for each sub-section and produces a variable total cognitive load (CL) represented by the solid line. Finally, the students' cognitive capacity, although necessarily variable, is illustrated as a constant by the crossed solid line.

In this case, we would anticipate sub-sections 7 and 15 would induce a cognitive overload. The student would most likely review these visually dense sub-sections or slow down the audio narration thus lowering the total cognitive load to capacity level. This reduction in WPM count would produce the effect shown in Figure 7.2.

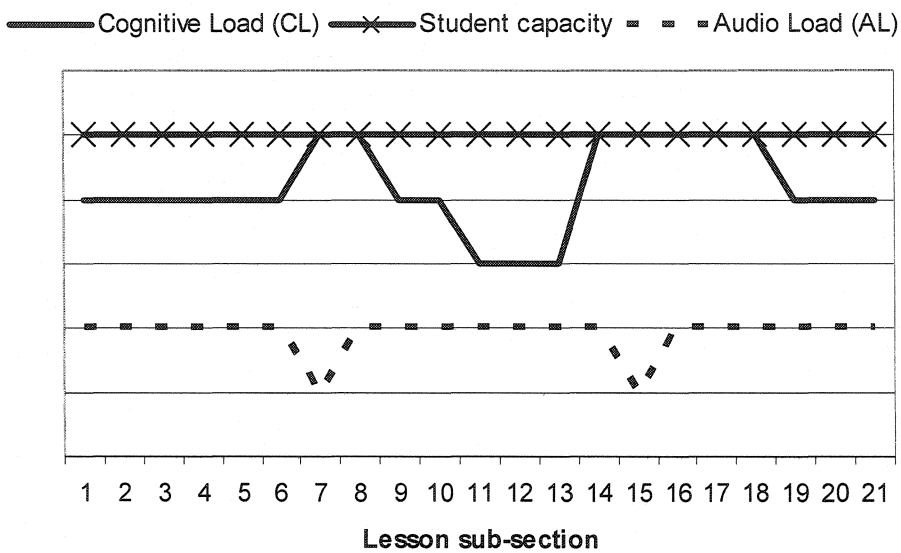

Figure 7.2. Cognitive load modified by speech expansion usage

Thus, with the ability to slow down the audio narration, the student is able to fend off cognitive overload and possibly save a rewind. Adversely, when the cognitive load demanded by the lesson is below capacity, in order to fend off gaps in concentration, an acceleration of the narration is recommended. Ultimately, the complete lesson can be dynamically paced to match ones' ability to ingest the material as illustrated in Figure 7.3.

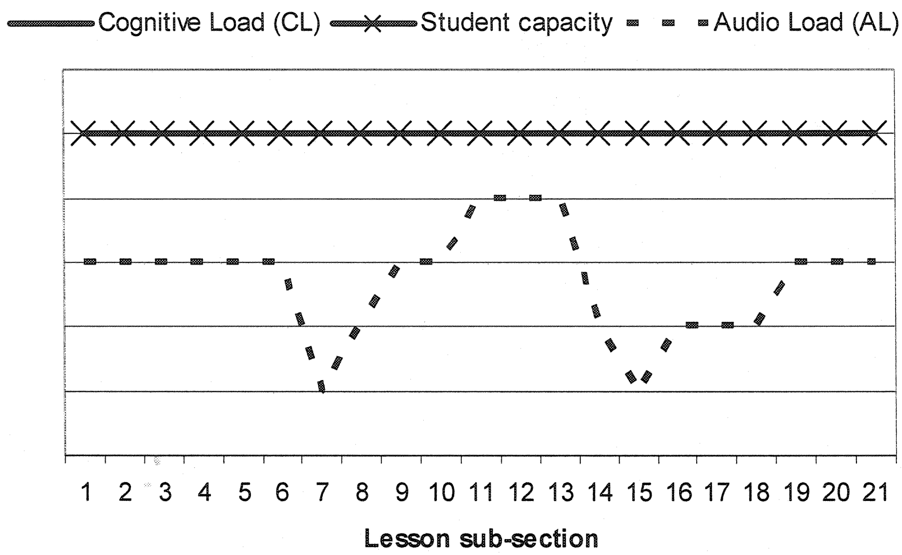

Figure 7.3. Cognitive load levelled to capacity using TSM

The cognitive load induced by the audio narration is increased and lowered to palliate the variable visual load demand and the resulting total cognitive load is merged and identical to the student's capacity. Although precise matching is not feasible, we believe dynamically slowing down or speeding up the audio narration were indeed attempts by the students to correspond lesson cognitive demand with their individual capacity.

7.4. Educative TSM acceptance

Throughout this research on time-altered speech and related video compression, we have asked ourselves the obvious question:

"If we've known this works since the '60s, how come it's not widely available?"

The answer has proved most difficult to elaborate. The knowledge gained through this research allows us to present facts which shed some light onto this lack of TSM availability. These can be grouped under the historical, economic, scientific and socio-cultural contexts.

7.4.1. Historical Context

In the 1960's, after the initial TSM development and scientific support, electronic devices were developed and marketed allowing time-altering of audio recorded on magnetic tape. Although the product delivered a quality signal, this mass distribution attempt was a commercial failure. The costs involved in purchasing the stand alone electronic apparatus were high and the consumer had no prior knowledge of TSM. As a promoter, one had to first educate concerning the technology and its benefits, equate this value with a high up front cost and naturally, find people who consumed high amounts of audio narration. A hard sell at best.

When the personal computer era came to be in the 1980's, processing requirements TSM algorithms demanded were beyond chip capacity and rendered real-time usage impossible. As processing power amazingly progressed according to Moore's law, the mid 1990's offered PCs with adequate chip capacity. Thus, thirty years had to pass for technical hurdles to be overcome and costs are now reduced to a 35$ software purchase in the case of the Enounce plug-in for example.

Unfortunately, TSM is not offered as an integrated feature within online educational course development software. In a typical project, once content is defined by the domain expert and a storyboard is produced, developers are expected to quickly produce the online material using timeline metaphor software such as Macromedia Director for example. These types of products necessarily offer a finite set of interface widgets which can be offered to the online student. Since such a TSM widget is not currently available, offering this functionality would force the educational content provider to foot the development costs of producing a third party plug-in.

Finally, the PC as an educational tool is slowly maturing. Although much anecdotal data is available, only Richard E. Mayer has developed a set of online educational design guidelines based on cognitive theory and subsequent scientific experimentation. Generally, educational technology applications are spearheaded by individual initiatives. For example, the few applications existing at l'École Polytechnique of Montreal have at their base, the course teacher as primary champion. This creates an effect where online content is disparate and more akin to an art form; each case is unique.

Thus, promotion of TSM in education does not have a clear champion. Educators are rarely technologically proficient and are trained in text based pedagogical design. Adversely, technical professionals do not have credibility in the soft human world of education where technology is not believed to have much affects on learning. Institutions with a tradition of didactics might produce such promoters who have the goal of producing tools to help learning.

7.4.2. Economic Context

Business and marketing experts may point out the problem is simply not technical but resides in the absence of a practical business model. There are no clear answers as to whom should pay for TSM availability, when and how much. In the BYU[1] deployment (Galbraith and Spencer 2001) for example, in the vast majority of cases the individual student purchased the Enounce TSM plug-in with personal funds. Most educational institutions would not accept accessibility based on financial capacity.

Current embryonic demand creates an additional commercialization hindrance. Although we are not marketing experts, creating a market for a specific technology inevitably passes through education. In this case, we believe promising pilot projects performed by innovative institutions wishing to position educational technology as a competitive advantage would bolster this marketing drive.

Lastly, the costs involved in producing audio narrated online material are significantly higher as compared with the written text alternative. Indeed, professional voices and high quality audio recording capabilities are not cheap. This generates a foreseeable affect where online content project managers will first budget and develop a written text design. Subsequent development phases may include audio narration if funding is available. Considering budget restraints in education, we have often encountered the notion that a text based online lesson

[1] Brigham Young University in Provo, Utah

might not be as good as an audio narrated version but it is still better than nothing. We would argue that in the majority of cases it is simply a digitally delivered paper based lesson and the computer is reduced to the role of a high tech mailman.

7.4.3. Scientific Context

Due to negative publicity brought about by large expenditures with little or no measurable effects on learning and unrealistic commercial hype, educational technology has lost mass-media interest. Scientific research and development projects in this area are thus difficult to fund and results are inevitably criticized since pinpointing and controlling all variables influencing human knowledge acquisition is at best an arduous task. Indeed, learning theories have radically changed and evolved in the last century and the human mind is still largely a mystery.

Additionally, TSM is not a new discovery and subsequently generates little fanfare related to hot new technologies. Informed readers know how it works, effects (or lack of) on learning effectiveness have been repeatedly published and efficiency gains have been reported. Empirical studies show a strong user preference for additional control over the learning process which is supported by pedagogical and human-computer interaction theories. There plainly seems to be little hard science left to be done.

Lastly, with the exception of the PC games industry, audio is generally underutilized in computing. Consequently, cognitive effects are little understood and few researchers have showed interest on the subject of sound usage in computing. Indeed, since audio is used in a multimodal context it becomes difficult to isolate its direct affects from interactions with the visual channel and more research is required in this area.

7.4.4. Socio-cultural Context

Our Western society is based on the written word as opposed to, for example, the African oral culture. As a result, audio narration demand is confined to niche markets and students expect learning involves reading. Indeed, from a very early age, social pressures and the public educational system enforce and promote reading abilities at the expense of listening skills. We are a society of readers and factual meaning as opposed to listening and interpretive multimodal communication. In this context, it is not surprising sound and audio narration in computing is not ubiquitous.

Additionally, higher education heavily favours strong readers as opposed to listeners. These same educated adults become domain experts, content developers, pedagogical and educational specialists. We believe it is reasonable to assume a strong written text bias is present within the complete chain of online content development actors. How would online content design differ if led by oral story tellers? We expect it would differ from a text based design and the related audio narrated lesson quality would be boosted.

As a whole, we believe the factors grouped within the above contexts significantly support understanding of the complex phenomenon which is the dissemination and mass utilization of technology. Although this list is not considered exhaustive it has satisfied our curiosity concerning TSM absence in current online training offering and we hope it may do the same for the reader.

7.5. Future research

Although this research concentrates on TSM and related visual content compression, this subject falls within the greater domain of skimming and searching through multimedia information. As proficient readers have developed methods to efficiently scan and evaluate great amount of printed text, "an interesting future direction is technology-assisted tools that allow authors to very quickly indicate important segments. (e.g., speech to-text transcript marked by author in 5 minutes using a tool)" (He, Sanocki, Gupta and Grudin 2000, p.184).

This would allow one to quickly 'scan' greater amounts of audio-visual information supported by skimming and navigation tools demonstrated by the SpeechSkimmer application presented in (Arons 1997) for example.

CONCLUSION

This memoir initially led us through a literary review of multimedia usage in education. This domain is still in its infancy and is mainly comprised of context specific case studies and anecdotes. Although most believe the digital media has a place in education the definition of its role is not yet clear. Opposing those who believe computing and the internet will provide universally accessible education we find others who believe learning is simply not influenced multimedia.

The digital medium's audio/visual and interactive capabilities were presented as the main edge over all other educational media (written word, television or audio only). They comprise the main reasons why a student should choose an online tutorial as an effective and efficient learning support tool.

Cognitive load theory and much experimental research by Richard E. Mayer allows one to state an audio narrated lesson augmented by static or dynamic visual information is cognitively more efficient than the equivalent printed narration. We have seen that reading is in fact an internal audio narration first interpreted by the visual channel. Providing an audio narration frees the visual channel of printed character interpretation and reduces risks of cognitive overload since both channels possess parallel short term memories and interpretation capabilities.

Like written text, an audio narrated lesson should allow individual dynamic pace control of content delivery. Indeed, a reader often slows down or speeds up the number of words per minute read in accordance to a multitude of factors such as prior domain knowledge and actual cognitive capacity. This type of behaviour should be possible for the online student using a pre-recorded audio/visual lesson.

Indeed, as explained by David H. Reilly, pace of content delivery is believed to be critical for efficient knowledge acquisition. A teacher in a typical classroom setting instinctively uses a rate of presentation which is perceived neither too slow nor too fast. Alternatively, some in the group will inevitably reach their personal threshold and find the rhythm too fast while others will be inadequately stimulated. Because of the one to one relationship between the computer and the user, an e-learner can functionally receive content at his or her own optimal pace.

We have presented time-scale modification technology as a way to provide the e-learner with the ability to dynamically modify the rate of speech and related visual information. This technology has existed for almost 40 years and effectively shortens a speech signal without modifying the pitch. The result is an audio narration whose words are heard being uttered proportionally quicker and still fully comprehensible. Although many TSM algorithms are available, current offering is based on overlap and add (OLA) principles producing a high quality time-scale modified signal.

Usage of TSM is education has been researched both experimentally and empirically. Scientific literature heralds efficiency gains and user appreciation. Two recent empirical studies, one at Brigham Young University and the other at Microsoft Corporation, have shown TSM can be beneficial to the consumption of digitally pre-recorded audio/visual lectures. Even content judged technical in nature is said to be more efficiently transmitted with the aid of TSM.

With such fanfare, we wondered why this technology was not universally available. It would seem educators do not see technology as being relevant to pedagogical design and technologists have difficulty promoting their tools for the art of teaching. This type of situation can only be resolved through repeated empirical evidence of benefits and scientific support.

The latter is found in the scientific literature but we found there seemed to be no empirical studies concerning TSM use with technical online tutorials as opposed to pre-recorded lectures. We expected a technical tutorial to offer less possibility of efficiency gain and ventured to guess some students might even slow down the pace of delivery below normal.

An empirical study was performed with the undergraduate engineered materials course at l'École Polytechnique de Montréal. The students were given access to a special version of the first chapter of the CD-ROM based course tutorial allowing dynamic rate modification through TSM. They were then asked to fill a survey and, five subjects were observed using the functionality.

The collected surveys were statistically analysed and, as expected, revealed a marked efficiency gain and a strong TSM user support. The average pace most used was 1.3 and 90% of subjects rated the functionality's usefulness with a score of 4 or more on a scale of 1 to 7. Some subjects reported TSM had little value but most asserted the increased control over the learning experience was appreciated.

Indeed 70% of respondents asserted dynamically varying the pace of instruction according to content complexity and only 17% stated they kept the same speed from beginning to end. In light of these numbers we applied a caveat to the reported 30% gain in efficiency since the average of most used speeds may not adequately represent actual average pace of instruction when it is dynamically changed throughout the lesson.

We believe the extensive use of the TSM functionality by a majority of the students illustrates the varying level of information density and complexity through the half hour lesson and consequently, if the same normal pace is imposed some segments will be watched repeatedly while others will bore the learner. The user comments and survey statistics lead us to state the main benefit TSM offers is increased user control over the learning experience rather than time savings.

The live observations provided us with real life anecdotes and usage situations. Although the behaviours were very different, most interesting was the tendency to inversely modify the rate of delivery in proportion to the visual information density. Indeed, when a tutorial segment showed strictly a video, speeds were highest. Alternatively, when the segment presented a mathematical demonstration along with an animated 3D model of the phenomenon, speed was reduced.

This suggests material with relatively constant visual information density like a pre-recorded lecture would allow students to set a relatively constant pace of delivery. Conversely, tutorials like the one used in this empirical study have highly variable visual density segments ranging from general description to mathematical demonstrations and related 3D animations. This variability would make the optimal pace of instruction for a particular student highly dependant on the visual density of the segment being watched.

We believe a strong case for TSM usage in education has been made. Wide spread offer is believed to be possible through promotion and the development of integrated TSM tools within online content editing software such as Macromedia Director.

REFERENCES

AARNTZEN, D. (1993). Audio in courseware - design knowledge issues. *Educational & Training Technology International*. 30:4. 354-366.

AGGARWAL, Anil. (2003). *Web-based education:learning from experience*. 1st ed. Hershey, PA: Information Science Publishing. 469p.

ALESSI, Stephen M., TROLLIP, Stanley R. (2001). *Multimedia for learning*. 3rd ed. Needham Heights, MA: Allyn and Bacon. 580p.

ARONS, Barry. (1991). Hyperspeech: Navigating in speech-only hypermedia. *Proceedings of Hypertext '91*. ACM. 133-146.

ARONS, Barry. (1997). SpeechSkimmer: A system for interactively skimming recorded speech. In *MIT Media Lab* site. [On line]. http://xenia.media.mit.edu/~barons/tochi97.html (Accessed May 12 ,2003)

BAEKER, Ronald, SMALL, Ian. (1990) *Animation at the Interface*. In The Art of human-computer interface design. LAUREL, Brenda. 1st ed. Reading, MA: Addison-Wesley Publishing Company. P. 251-267.

BARRON, Ann E., KYSILKA, Marcella L. (1993). The effectiveness of digital audio in computer-based training. *Journal of Research in Computing in Education*. 25:3. 277-289.

BEDGAR, Anne K., CUNNINGHAM, Donald J., DUFFY, Thomas M., PERRY, J. David. (1992) *Theory into practice: how do we link?* In Constructivism and the Technology of Instruction: a conversation. DUFFY, Thomas M., JONASSEN, David H. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates. P. 17-34.

BERG, Gary A. (2003). *The knowledge medium.* 1st ed. Hershey, PA: Information Science Publishing. 264p.

BEUSCHEL, Werner, GAISER, Birgit, DRAHEIM, Susanne. (2003) *Communication needs of online students.* In Web-based education:learning from experience. AGGARWAL, Anil. 1st ed. Hershey, PA: Information Science Publishing. P. 203-221.

BISHOP, M. J., CATES, Ward Mitchell. (2001). Theoretical foundations for sound's use in multimedia instruction to enhance learning. *Educational Technology Research and Development.* 49:3. 5-22.

BLOOM, B. (1969). *Taxonomie des objectifs d'apprentissage. Domaine cognitif. Tome1.* Montréal: Éditions Nouvelles. 232p.

BOYLE, Tom. (2002). Towards a theoretical base for educational multimedia design. *Journal of Interactive Media in Education.* [On Line]. 2002:2. www-jime.open.ac.uk/2002/2/index.html (Accessed July 19, 1903)

BRUCE, Betsy. (2002). *eLearning with Dreamweaver MX : building online learning applications.* 1st ed. Macromedia Press, New Riders Publishing. 564p.

BRUNING, Roger, HORN, Christy A., PYTLIKZILLIG, Lisa M. (2003). *Web-based learning: What do we know? Where do we go?* 1st ed. Greenwich, CT: Information Age Publishing. 275p.

BURKE, Robin D. (1998) *Representation, storage, and retrieval of tutorial stories is a social simulation.* In Inside Multi-Media Case Based Instruction. SCHANK, Roger C. 1st ed. Mahwah, NJ: Lawrence Erlbaum Associates. P. 175-284.

CAMACHO, Cynthia, FOULDS, Richard. (2002). Enhancing access to IT for persons with disabilities. *Proceedings of the IEEE 28th annual Northeast Bioengineering Conference*. IEEE. 141-142.

CARD, Stuart K. (1989) *Models of Working Memory*. In Human performance models for computer-aided engineering. ELKIND, J. E., CARD, S. K., HUCHBERG, J., MESSICK HUEY, B. Washington, D.C.: National Academy Press. P. 203-214.

CERRI, Stefano A., GOUARDÈRES, Guy, PARAGUAÇU, Fabio. (2002). *Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2002, Proceedings*. Berlin, Germany: Springer.

CLARK, Richard E. (2001a) *Do media aid in critical thinking, problem solving or 'learning to learn'?* In Learning from media: arguments, analysis and evidence. CLARK, Richard E. 1st ed. Greenwich, CT: Information Age Pub. P. 103-124.

CLARK, Richard E. (2001b) *Media are 'mere vehicles' : the opening argument*. In Learning from media: arguments, analysis and evidence. CLARK, Richard E. 1st ed. Greenwich, CT: Information Age Pub. P. 103-124.

CLARK, Richard E. (2001c) *New directions: cognitive and motivational research issues*. In Learning from media: arguments, analysis and evidence. CLARK, Richard E. 1st ed. Greenwich, CT: Information Age Pub. P. 263-298.

CLARK, Richard E. (2001d) *What is next in the media and methods debate?* In Learning from media: arguments, analysis and evidence. CLARK, Richard E. 1st ed. Greenwich, CT: Information Age Pub. P. 327-337.

CLARK, Richard E. (2003) *Research on web-based learning : a half-full glass*. In Web-based learning: What do we know? Where do we go? BRUNING, Roger, HORN, Christy A., PYTLIKZILLIG, Lisa M. 1st ed. Greenwich, CT: Information Age Publishing. P. 1-22.

CLARK, Richard E., CRAIG, Terry. (2001) *What about multi-media effects on learning?* In Learning from media: arguments, analysis and evidence. CLARK, Richard E. 1st ed. Greenwich, CT: Information Age Pub. P. 89-101.

CLARK, Ruth Colvin, MAYER, Richard E. (2003). *E-Learning and the science of instruction : proven guidelines for consumers and designers of multimedia learning*. 1st ed. San Francisco, CA: Jossey-Bass/Pfeiffer. 322p.

COGNITION AND TECHNOLOGY GROUP AT VANDERBUIL UNIVERSITY. (1992) *Technology and the design of generative learning environments*. In Constructivism and the Technology of Instruction: a conversation. DUFFY, Thomas M., JONASSEN, David H. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates. P. 77-89.

COGNITION AND TECHNOLOGY GROUP AT VANDERBUIL UNIVERSITY. (2003) *Connecting learning theory and instructional practice: leveraging some powerful afforfances of technology*. In Technology applications in education. O'NEIL, Harold F. Jr., PEREZ, Ray S. 1st ed. Mahwah, NJ: Lawrence Erlbaum Associates. P. 173-209.

COLLINS, Rives, COOPER, Pamela J. (1997). *The Power of Story : Teaching through Storytelling*. 2nd ed. Scottsdale, AZ: Gorsuch Scarisbrick. 182p.

COVELL, M., WITHGOTT, M., SLANEY, M. (1998). Mach1: Non uniform Time-Scale Modification of Speech. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE.

CRAWFORD, Chris. (1990) *Lessons from computer game design*. In The Art of human-computer interface design. LAUREL, Brenda. 1st ed. Reading, MA: Addison-Wesley Publishing Company. P. 103-111.

CUNNINGHAM, Donald J. (1992) *Assessing construction and constructing assessments: a dialog*. In Constructivism and the Technology of Instruction: a conversation. DUFFY, Thomas M., JONASSEN, David H. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates. P. 35-44.

DA GRAÇA PIMENTEL, M., ISHIGURO, Y., KERIMBAEV, B., ABOWD, G. D., GUZDIAL, M. (2001). Supporting educational activities through dynamic web interfaces. *Interacting with Computers*. 13 353-374.

DESSUS, Philippe, LEMAIRE, Benoît. (2002). Using production to assess learning : an ILE that fosters self-regulated learning. *Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2002, Proceedings*. Berlin, Germany: Springer. 772-781.

DHOLAKIA, Ruby Roy, ZHAO, Miao, DHOLAKIA, Nikhilesh, FORTIN, David R. (2000). Interactivity and revisits to websites: a theoretical framework. In *RITIM* site. [On line]. http://ritim/cba/uri/edu/wp/ (Accessed Jun 18 ,2003)

DILLENBOURG, Pierre. (1999). *Collaborative learning : cognitive and computational approaches*. 1st ed. Oxford, UK: Pergamon. 246p.

DON, Abbe. (1990) *Narrative at the interface*. In The Art of human-computer interface design. LAUREL, Brenda. 1st ed. Reading, MA: Addison-Wesley Publishing Company. P. 383-391.

DUBOIS, Michel, VIAL, I. (2000). Multimedia design: the effects of relating multimodal information. *Journal of Computer Assisted Learning*. 16 157-165.

DUFFY, Thomas M., JONASSEN, David H. (1992). *Constructivism and the Technology of Instruction: a conversation*. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates. 221p.

DUKES, Darren R. (1995). *Real time time-scale modifications of speech (Unpublished).* 76p. University of Waterloo.

DUPOUX, Emmanuel, GREEN, Kerry. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception & Performance*. 23:3. 914-927.

DUTTON, William H., LOADER, Brian D. (2002). *Digital academe: the new media and institutions of higher education*. 1st ed. London, UK: Routledge. 368p.

EDELSON, Daniel C. (1998) *Learning from stories: an architecture for Socratic case based teaching*. In Inside Multi-Media Case Based Instruction. SCHANK, Roger C. 1st ed. Mahwah, NJ: Lawrence Erlbaum Associates. P. 103-174.

EISENSTADT, Marc, VINCENT, Tom. (2000). *The knowledge web: learning and collaborating on the net*. 1st ed. London, UK: Kogan Page Limited. 295p.

FAIRBANKS, G., EVERITT, W. L., JAEGER, R. P. (1954). Method for time or frequency compression-expansion of speech. *Transactions of the Institute of Radio Engineers, Professional Group on Audio*. AU2:1. 7-12.

FORCIER, Richard C. (2002). *The computer as an educational tool : productivity and problem solving*. 3rd ed. Upper Saddle River, NJ: Merrill/Prentice Hall. 446p.

FULFORD, Catherine P. (1992). Systematically designed text enhanced with compressed speech audio. *EDRS*.

GADE, Paul A., MILLS, Carol B. (1989). Listening rate and comprehension as a function of preference for and exposure to time-altered speech. *Perceptual & Motor Skills*. 68:2. 531-538.

GALBRAITH, Joel D., SPENCER, Steven G. (2001). *Variable speed playback of digitally recorded lectures: evaluating learner feedback.* 10p. Brigham Young University, Provo, Utah.

GERLIC, Ivan, JAUSOVEC, Norbert. (1999). Multimedia: differences in cognitive processes observed with EEG. *Educational Technology Research and Development*. 47:3. 5-14.

GHAOUI, Claude. (2003). *Usability Evaluation of Online Learning Programs*. 1st ed. Hershey, Pa: Information Science Publishing. 423p.

GRIFFIN, D. W., LIM, J. S. (1984). Signal estimation from modified short-time Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 32:2. 236-243.

HARA, Noriko, KLING, Rob. (2002) *Students' difficulties in a Web-based distance education course: an ethnographic study*. In Digital academe: the new media and institutions of higher education. DUTTON, William H., LOADER, Brian D. 1st ed. London, UK: Routledge. P. 62-84.

HARDAM, E. (1990). High quality time scale modification of speech signals using fast synchronized-overlap-add algorithms. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE. 409-412.

HARRIGAN, Kevin. (1995). The SPECIAL System: self-paced education with compressed interactive audio learning. *Journal of Research in Computing in Education*. 27:3. 361-370.

HARRIGAN, Kevin. (1996). *Just noticeable difference and effects on searching of user-controlled time-compressed digital-video.* 146p. PhD Thesis Department of Education, University of Toronto.

HARRIGAN, Kevin. (2000). The SPECIAL System: searching time-compressed digital video lectures. *Journal of Research in Computing in Education*. 33:1. 77-86.

HE, Liwei, GRUDIN, Jonathan, GUPTA, Anoop. (2000). Designing presentations for on-demand viewing. *Proceedings of ACM conference on Computer supported cooperative work 2000*. ACM. 127-134.

HE, Liwei, GUPTA, Anoop. (2001). Exploring benefits of non-linear time compression. *Proceedings of Multimedia 2001*. ACM. 382-391.

HE, Liwei, SANOCKI, Elizabeth, GUPTA, Anoop, GRUDIN, Jonathan. (2000). Comparing presentation summaries: Slides vs. Reading vs. Listening. *Proceedings of CHI 2000*. ACM. 2, 177-184.

HERGENHAHN, B. R. (1976). *An introduction to theories of learning*. 1st ed. Englewood Cliffs, NJ: Prentice-Hall, INC. 402p.

HILDYARD, A., OLSEN, D. R. (1982) *On the comprehension of oral vs written discourse*. In Spoken and written language:Exploring orality and literacy. TANNEN, D. Norwood, NJ: Ablex. P. 19-24.

HINDUS, Debby, SCHMANDT, Chris, HORNER, Chris. (1993). Capturing, structuring and representing ubiquitous audio. *ACM Transactions on Information Systems*. 11:4.

HORTON, William, HORTON, Katherine. (2003). *E-Learning tools and technologies*. 1st ed. Indianapolis, IN: Wiley Publishing INC. 574p.

IGO, L. Brent, BRUNING, Roger, MCCRUDDEN, Matthew, KAUFFIN, Douglas F. (2003) *INFOGATHER : a tool for gathering and organizing information from the web.* In Web-based learning: What do we know? Where do we go? BRUNING, Roger, HORN, Christy A., PYTLIKZILLIG, Lisa M. 1st ed. Greenwich, CT: Information Age Publishing. P. 57-77.

JASPER, Fons. (1991). The relationship of sound-image. *International Journal of Instructional Media*. 18:2. 161-174.

JONA, Menachem Y. (1998) *Representing and applying teaching strategies in computer-based learning-by-doing tutors*. In Inside Multi-Media Case Based Instruction. SCHANK, Roger C. 1st ed. Mahwah, NJ: Lawrence Erlbaum Associates. P. 285-443.

JONASSEN, David H., HOWLAND, Jane. (2003). *Learning to solve problems with technology : a constructivist perspective*. 2nd ed. Upper Saddle River, NJ: Merrill/Prentice Hall. 256p.

JONES, Miles R. (1986). The Effects of Training with Time Compressed Speech Plus Text on Reading Speed, Vocabulary, and Comprehension. *EDRS*.

JUNOR, Lloyd. (1992). Teaching by tape: Some benefits, problems and solutions. *Distance Education*. 13:1. 93-107.

KAROULIS, Athanasis, POMBORTSIS, Andreas. (2003) *Heuristic evaluation of Web-based ODL programs*. In Usability Evaluation of Online Learning Programs. GHAOUI, Claude. 1st ed. Hershey, Pa: Information Science Publishing. P. 88-109.

KIM, Jihie, GIL, Yolanda. (2002). Deriving sequisition principles from tutoring principles. *Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2002, Proceedings*. Berlin, Germany: Springer. 661-670.

KING, Paul E., BEHNKE, Ralph R. (1989). The effect of time-compressed speech on comprehensive, interpretive, and short-term listening. *Human Communication Research*. 15:3. 428-443.

KING, Stephen. (1978) *The lawnmower man*. In Night Shift. KING, Stephen. Garden City, N.Y.: Doubleday. P. 336.

KOBAYASHI, Minoru, SCHMANDT, Chris. (1997). Dynamic soundscape: mapping time to space for audio browsing. *Proceedings of CHI '97*. ACM. 194-201.

KORT, Barry, REILLY, Rob. (2002). An affective module for an intelligent tutoring system. *Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2002, Proceedings*. Berlin, Germany: Springer. 955-962.

KOZMA, Robert B. (2001) *Robert Kozma's counterpoint theory of 'learning with media'*. In Learning from media: arguments, analysis and evidence. CLARK, Richard E. 1st ed. Greenwich, CT: Information Age Pub. P. 137-178.

LAI, Shu-Ling. (2000). Influence of audio-visual presentations on learning abstract concepts. *International Journal of Instructional Media*. 27:2. 199-206.

LAROSE, François, KARSENTI, Thierry. (2002). *La place des TIC en formation initiale et continue*. Sherbrooke, Québec: Éditions du CRP. 238 pp.

LAUREL, Brenda. (1990). *The Art of human-computer interface design*. 1st ed. Reading, MA: Addison-Wesley Publishing Company. 522p.

LAUREL, Brenda. (1993). *Computers as theatre*. 1st ed. Reading, MA: Addison-Wesley Publishing Company. 227p.

LAURILLARD, Diana, STRATFOLD, M., LUCKIN, R., PLOWMAN, L., TAYLOR, Josie. (2000). Affordances for learning in a non-linear narrative medium. *Journal of Interactive Media in Education*. [On Line]. 2000:2. www-jime.open.ac.uk/2000/2/index.html (Accessed July 18, 1903)

LEE, S., KIM, H. (1997). Variable time-scale modifications of speech using transient information. *IEEE: Proceedings of the international conference on acoustics, speech, and signal processing*. Munich: IEEE. 2, 1319-1322.

LI, Francis C., GUPTA, Anoop, SANOCKI, Elizabeth, HE, Liwei, RUI, Yong. (2000). Browsing digital video. *Proceedings of CHI '2000*. 169-176.

LOCKWOOD, Fred, GOOLEY, Anne. (2001). *Innovation in open & distance learning*. 1st ed. London, UK: Kogan Page Limited. 227p.

MALONE, Thomas W., LEPPER, Mark R. (1987) *Making learning fun: a taxonomy of intrinsic motivations for learning*. In Cognitive and Affective Process Analysis : volume 3 of Aptitude. SNOW, Richard E., FARR, Marshall J. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates. P. 223-253.

MANN, Bruce L. (1995). Enhancing educational software with audio: assigning structural and functional attributes from the SSF Model. *British Journal of Educational Technology*. 26:1. 16-29.

MANN, Bruce L. (1997). Evaluation of presentation modalities in a hypermedia System. *Computers & Education* . 28:2. 133-143.

MAYER, Richard E. (2003a). *Learning and instruction*. 1st ed. Upper Saddle River, NJ: Prentice-Hall.

MAYER, Richard E. (2003b). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction*. 13:2. 125-139.

MAYER, Richard E. (2003c) *Theories of learning and their application to technology*. In Technology applications in education. O'NEIL, Harold F. Jr., PEREZ, Ray S. 1st ed. Mahwah, NJ: Lawrence Erlbaum Associates. P. 127-157.

MAYER, Richard E., CHANDLER, P. (2001). When learning is just a click away: does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*. 93 390-397.

MAYER, Richard E., MORENO, Roxanna. (2003) *Nine ways to reduce cognitive load in multimedia learning*. In Web-based learning: What do we know? Where do we go? BRUNING, Roger, HORN, Christy A., PYTLIKZILLIG, Lisa M. 1st ed. Greenwich, CT: Information Age Publishing. P. 23-44.

MAYES, J. T., FOWLER, C. J. (1999). Learning technology and usability: a framework for understanding courseware. *Interacting with Computers*. 11 485-497.

MORKES, J., NIELSON, J. (1997). Concise, scannable, and objective: How to write for the web. www.useit.com/papers/webwriting/writing.html (Accessed Jan 19 ,2004)

MOUSAVI, Seyed Yaghoub, LOW, Renae, SWELLER, John. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*. 87:2. 319-334.

MUILENBURG, Lin, BERGE, Zane L. (2000). A framework for designing questions for online learning. *DEOSNEWS* . [On Line]. 10:2. http://www.ed.psu.edu/acsde/deos/doesnews/deosnews10_2.asp (Accessed Aug. 31, 1903)

MURRAY, Janet H. (1997). *Hamlet on the Holodeck : the future of narrative in cyberspace*. 1st ed. New York, NY: The Free Press. 324p.

NEGROPONTE, Nicholas. (1995). *being digital*. 1st ed. New York, NY: Alfred A. Knopf. 243p.

NICOL, Anne. (1990) *Interfaces for learning: what do good teachers know that we don't?* In The Art of human-computer interface design. LAUREL, Brenda. 1st ed. Reading, MA: Addison-Wesley Publishing Company. P. 113-122.

NORMAN, Donald A. (1982). *Learning and memory*. 1st ed. San Francisco, CA: W.H. Freeman and Company. 129p.

NORMAN, Donald A. (1988). *The psychology of everyday things*. 1st ed. New York, NY: Basic Books INC. 257p.

O'NEIL, Harold F. Jr., PEREZ, Ray S. (2003). *Technology applications in education*. 1st ed. Mahwah, NJ: Lawrence Erlbaum Associates. 372p.

OLSON, Janet S. (1985). A Study of the relative effectiveness of verbal and visual augmentation of rate-modified speech in the presentation of technical material. *EDRS*.

OLSON, Janet S., BERRY, Louis H. (1982). The state of the art in rate-modified speech: a review of contemporary research. *EDRS*.

OMOIGUI, N., HE, Liwei, GUPTA, Anoop, GRUDIN, Jonathan, SANOCKI, Elizabeth. (1999). Time-compression: System concerns, usage, and benefits. *Proceedings of ACM Conference on Computer Human Interaction*. ACM. 136-143.

OREN, Tim. (1990) *Designing a new medium*. In The Art of human-computer interface design. LAUREL, Brenda. 1st ed. Reading, MA: Addison-Wesley Publishing Company. P. 467-479.

PAAS, Fred, RENKL, Alexander, SWELLER, John. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist*. 38:1. 1-4.

PALLOF, Rena M. (2003). *The virtual student: a profile and guide to working with online learners*. 1st ed. San Francisco, CA: Jossey-Bass. 191p.

PAQUETTE, Gilbert. (2002). *L'ingénierie pédagogique*. 1st ed. Sainte-Foy, QC: Presses de l'Université du Québec. 456p.

PARLENGELI, Oronzo, MARCHIGIANI, enrica, BAGNARA, Sebastiano. (1999). Multimedia systems in distance education: effects of usability on learning. *Interacting with Computers*. 12 37-49.

PARSHALL, Cynthia G. (1999). Audio CBTs: measuring more through the use of speech and non-speech sound. *EDRS*.

PAUL, Sarla. (1992). Pictorial Redundancy, Processing Time, Number of Propositions, and Comprehension of Compressed Speech. *International Journal of Instructional Media*. 19:1. 71-78.

PERKINS, David N. (1992a) *Technology meets constructivism: do they make a marriage?* In Constructivism and the Technology of Instruction: a conversation. DUFFY, Thomas M., JONASSEN, David H. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates. P. 57-76.

PERKINS, David N. (1992b) *What constructivism demands of the learner*. In Constructivism and the Technology of Instruction: a conversation. DUFFY, Thomas M., JONASSEN, David H. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates. P. 161-165.

PHILLIPS, D. C. (1995). The good, the bad, the ugly: the many faces of constructivism. *Educational Researcher*. 24:7. 5-12.

PLOETZER, Rolf, DILLENBOURG, Pierre, PREIER, Michael, TRAUM, David. (1999) *Learning by explaining to oneself and to others*. In Collaborative learning : cognitive and computational approaches. DILLENBOURG, Pierre. 1st ed. Oxford, UK: Pergamon. P. 103-121.

REILLY, David H. (1998). The pace of instructional presentation, non-linear systems, effectiveness of cognitive processing, and needed research in the use of technology for instruction. *Journal of Instructional Psychology*. 25:3. 151-165.

RICKEL, Jeff, LESH, Neal, RICH, Charles, SIDNER, Candace L., GERTNER, Abigail. (2002). Collaborative discourse theory as a foundation for tutorial dialogue. *Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2002, Proceedings*. Berlin, Germany: Springer. 542-551.

ROUANE, Khalid, FRASSON, Claude, KALTENBACH, Marc. (2002). LKC : Learning by Knowledge Construction. *Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2002, Proceedings*. Berlin, Germany: Springer. 188-198.

ROUCOS, S., WILGUS, A. M. (1985). High-quality time-scale modifications for speech. *IEEE: Proceedings of the international conference on acoustics, speech, and signal processing*. IEEE. 493-496.

RUMELHART, David E., NORMAN, Donald A. (1978) *Accretion, tuning and restructuring: three modes of learning*. In Semantic factors in cognition. COTTON, J. W., KLATZKY, R. L. New Jersey: Hillsdale. P. 37-53.

RYAN, Marie-Laure. (2001). *Narrative as Virtual Reality : Immersion and Interactivity in Literature and Electronic Media*. 1 ed. Baltimore, MD: The Johns Hopkins University Press. 399p.

SALMON, Gilly. (2002). *etivities: the key to active online learning*. 1st ed. London, UK: Kogan Page Limited. 223p.

SANDERS, William B. (2001). *Creating learning-centered courses for the world wide web*. 1 ed. Boston, MA: Allyn and Bacon. 302p.

SCHANK, Roger C. (1998). *Inside Multi-Media Case Based Instruction*. 1 ed. Mahwah, NJ: Lawrence Erlbaum Associates. 451p.

SCHWAB, Eileen C., DEGROOT, Jenny. (1993). Listener response to time compressed speech. *Proceedings of CHI 93*. ACM. 81-82.

SCOTT, Peter, PHILLIPS, Marion. (2000) *Developing Web-based student support systems: telling the student stories on the internet*. In The knowledge web : learning and collaborating on the net. EISENSTADT, Marc, VINCENT, Tom. 1 ed. London, UK: Kogan Page Limited. P. 79-95.

SIEBER, Vivien, ANDREW, David. (2003) *Learning technologies and learning theories*. In Usability Evaluation of Online Learning Programs. GHAOUI, Claude. 1 ed. Hershey, Pa: Information Science Publishing. P. 218-232.

SINCLAIR, Joseph T. (2002). *Creating Web-based training : a step-by-step guide to designing effective E-learning*. 1 ed. New York, NY: Amacom. 582p.

SNOW, Richard E., FARR, Marshall J. (1987). *Cognitive and Affective Process Analysis : volume 3 of Aptitude*. 1 ed. Hillsdale, NJ: Lawrence Erlbaum Associates. 370p.

SQUIRES, D., PREECE, J. (1999). *Predicting quality in educational software: evaluating for learning, usability and synergy between them. Interacting with computers.*

STIFELMAN, Lisa, ARONS, Barry, SCHMANDT, Chris. (2001). The audio notebook. *Proceedings of CHI 2001*. ACM. 3, 182-189.

STONEY, Sue, OLIVER, Ron. (1999). Can higher order thinking and cognitive engagement be enhanced with multimedia? *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning* . [On Line]. imej.wfu.edu/articles/1999/2/07/index.asp (Accessed Aug. 27, 1903)

SUCHMAN, Lucy A. (1987). *Plans and situated actions the problem of human-machine communication*. Cambridge Cambridgeshire, New York: Cambridge University Press. xii, 203 pp.

SUMNER, Tamara, TAYLOR, Josie. (2000) *Media integration through meta-learning environments*. In The knowledge web : learning and collaborating on the net. EISENSTADT, Marc, VINCENT, Tom. 1 ed. London, UK: Kogan Page Limited. P. 63-78.

SWELLER, John. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

THORSEN, Carolyn. (2003). *TechTactics : instructional models for educational computing*. 1 ed. Boston, MA: Allyn and Bacon. 281p.

VERHELST, Werner, ROELANDS, Marc. (1993). An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE. 554-557.

VICENTE, Angel de, PAIN, Helen. (2002). Informing the detection of the students' motivational state : an empirical study. *Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2002, Proceedings*. Berlin, Germany: Springer. 933-943.

VYGOTSKY, L. S. (1978). *Mind in society*. 1st ed. Cambridge, MA: Harvard University Press.

WELLNER, Pierre D., WEIMER, David M., ARONS, Barry. (2001). Conference Scribe: Turning conference calls into documents. *Proceedings of the 34th Hawaii International Conference on System Sciences - 2001*. IEEE. 1609-1617.

APPENDICES

APPENDICE 1. STUDENT SURVEY

Ce questionnaire sert à évaluer l'utilité de la technologie de compression audio-visuelle TSM (Time-scale modification) dans un contexte éducationnel. Une fois que vous aurez complété le chapitre 1 du didacticiel « Des Matériaux » et utilisé le logiciel qui permet de modifier dynamiquement le débit de la narration, veuillez répondre aux questions qui suivent. Votre participation est volontaire, confidentielle et anonyme.

1. Sur une échelle de 1 à 7, encerclez le chiffre correspondant, selon vous, à l'utilité de la fonction d'accélération/décélération

1 2 3 4 5 6 7

Inutile Moyen Très utile

2. Quelle vitesse avez-vous utilisée le plus souvent (par ex.: 1.3 ou 0.8 ou 1.6)?

3. Encerclez **les** options qui décrivent le mieux votre utilisation (plus d'un choix possible).

a) Après une période d'adaptation, j'ai gardé la même vitesse

b) J'ai changé de vitesse selon la difficulté de la matière

c) J'ai écouté au moins un segment à une vitesse plus lente que normale

d) J'ai trouvé la vitesse qui me convient dès le début de la leçon

e) J'ai revu au moins un passage plus d'une fois, si oui, combien de fois (par ex. 2 ou 3 fois)?

______

f) J'ai utilisé la fonction d'accélération pour trouver un passage particulier

4. Sur une échelle de 1 à 7, encerclez le chiffre correspondant, selon vous, à la difficulté de la matière.

1 2 3 4 5 6 7

Facile Moyen Très difficile

5. Décrivez brièvement vos impressions de cette fonctionnalité.

Le sondage est terminé. Nous vous demandons de le remettre au professeur du cours ING1035 le plus tôt possible.

Nous vous remercions de votre participation à cette recherche. Nous espérons que les résultats permettront d'avancer les connaissances en technologies éducatives et, par le fait même, améliorer l'efficacité de l'apprentissage supporté par informatique.

Pour toutes questions ou commentaires n'hésitez pas à contacter :

Charles-Antoine Julien (charles-antoine.julien@polymtl.ca)

APPENDICE 2. COMPLETE SURVEY RESULTS

	Question			
Survey	1	2	3	4
1				5
2	5	1.7	b	4
3	4	1.7	a	5
4	3	1.3	d,e(3),f	4
5	7	1.2	b,c,e(3)	4
6	4	1	a	4
7	6	1.3	b,e(?),f	6
8	4	1.4	b	4
9	7	1.3	b	4
10	6	1.3	b,c	4
11	4	1.2	b,c,e(2),f	5
12	3	1	b,c,e(2)	4
13	4	?	b,c,e(2),f	4
14	6	1	c,d	5
15	4	1.3	b,c,d	6
16	6	1.2	b,c	6
17	5	1.5	b,f	6
18	7	1.2	a,b,e	1
19	4	1.4	b	5
20	5	?	b,e(?)	5
21	5	1.1	a,b,c,e(2)	2
22	4	1.3	d,e(3),f	4
23	7	1.3	b	4
24	6	1.3	b	4
25	4	?	d,f	5
26	6	1.2	d	6
27	5	1.3	b,c,f	2
28	7	0.9	b	6
29	4	1.3	a	5
30	7	1.6	b,e(2),f	6